

WEST Search History

DATE: Friday, January 03, 2003

<u>Set Name</u> side by side	<u>Query</u>	<u>Hit Count</u>	<u>Set Name</u> result set
<i>DB=USPT,PGPB,EPAB,DWPI; PLUR=YES; OP=OR</i>			
L26	l23 and paclitaxel	1	L26
L25	l23 and paclitaxol	0	L25
L24	L23 and taxol	2	L24
L23	peptide adj internaliz\$	33	L23
<i>DB=USPT,PGPB; PLUR=YES; OP=OR</i>			
L22	peptide adj internaliz\$	32	L22
L21	L20 and taxol	0	L21
L20	L19 and internaliz\$	1	L20
L19	5087616.pn.	1	L19
L18	L17 and internalize	0	L18
L17	5753204.pn.	1	L17
<i>DB=DWPI; PLUR=YES; OP=OR</i>			
L16	WO008800837a2	0	L16
<i>DB=EPAB; PLUR=YES; OP=OR</i>			
L15	WO008800837a2	0	L15
L14	WO008800837a1	0	L14
<i>DB=DWPI; PLUR=YES; OP=OR</i>			
L13	WO008800837a1	0	L13
L12	WO888800837a1	0	L12
L11	WO888800837A2	0	L11
<i>DB=EPAB; PLUR=YES; OP=OR</i>			
L10	L9	0	L10
<i>DB=USPT,PGPB,EPAB,DWPI; PLUR=YES; OP=OR</i>			
L9	WO888800837A2	0	L9
<i>DB=USPT,PGPB; PLUR=YES; OP=OR</i>			
L8	WO888800837A2	0	L8
L7	L6 and internalize	1	L7
L6	L4 and tumor	1	L6
L5	L4 and peptide adj internaliz\$	0	L5
L4	5847121.pn.	1	L4
L3	tumor cell and L2	59243	L3
L2	L1 and internaliz\$	1	L2
L1	5587459.pn.	1	L1

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 14:27:08 ; Search time 10 Seconds
(without alignments)
22.742 Million cell updates/sec

Title: US-09-899-376-1
Perfect score: 12
Sequence: 1 TSPINIHNGQKL 12

Scoring table: OLIGO
Gapext 60.0 , Gapext 60.0

Searched: 117078 seqs, 18951520 residues

Word size : 0

Total number of hits satisfying chosen parameters: 117078

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/us08_NEW_PUB_pep: *
2: /cgn2_6/ptodata/1/pubpaa/PCN_NEW_PUB_pep: *
3: /cgn2_6/ptodata/1/pubpaa/us06_NEW_PUBCOMB_pep: *
4: /cgn2_6/ptodata/1/pubpaa/us07_NEW_PUB_pep: *
5: /cgn2_6/ptodata/1/pubpaa/us07_PUBCOMB_pep: *
6: /cgn2_6/ptodata/1/pubpaa/us07_PUBCOMB_pep: *
7: /cgn2_6/ptodata/1/pubpaa/PCNTS_PUBCOMB_pep: *
8: /cgn2_6/ptodata/1/pubpaa/us08_PUBCOMB_pep: *
9: /cgn2_6/ptodata/1/pubpaa/us09_NEW_PUB_pep: *
10: /cgn2_6/ptodata/1/pubpaa/us09_PUBCOMB_pep: *
11: /cgn2_6/ptodata/1/pubpaa/us10_NEW_PUB_pep: *
12: /cgn2_6/ptodata/1/pubpaa/us10_PUBCOMB_pep: *
13: /cgn2_6/ptodata/1/pubpaa/us60_NEW_PUB_pep: *
14: /cgn2_6/ptodata/1/pubpaa/us60_PUBCOMB_pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/us08_NEW_PUB_pep: *
2: /cgn2_6/ptodata/1/pubpaa/PCN_NEW_PUB_pep: *
3: /cgn2_6/ptodata/1/pubpaa/us06_NEW_PUBCOMB_pep: *
4: /cgn2_6/ptodata/1/pubpaa/us07_NEW_PUB_pep: *
5: /cgn2_6/ptodata/1/pubpaa/us07_PUBCOMB_pep: *
6: /cgn2_6/ptodata/1/pubpaa/us07_PUBCOMB_pep: *
7: /cgn2_6/ptodata/1/pubpaa/us08_PUBCOMB_pep: *
8: /cgn2_6/ptodata/1/pubpaa/us09_NEW_PUB_pep: *
9: /cgn2_6/ptodata/1/pubpaa/us09_PUBCOMB_pep: *
10: /cgn2_6/ptodata/1/pubpaa/us10_NEW_PUB_pep: *
11: /cgn2_6/ptodata/1/pubpaa/us10_PUBCOMB_pep: *
12: /cgn2_6/ptodata/1/pubpaa/us60_NEW_PUB_pep: *
13: /cgn2_6/ptodata/1/pubpaa/us60_PUBCOMB_pep: *

RESULT 1
US-09-899-376-1

; Sequence 1, Application US/09899376
; Patent No. US20030102265A1
; GENERAL INFORMATION:
; APPLICANT: HONG, FRANK D.

; TITLE OF INVENTION: ISOLATION OF A CELL-SPECIFIC INTERNALIZING PEPTIDE THAT

; FILE REFERENCE: UTSC:645US

;

; CURRENT FILING DATE: 2001-07-02

;

;

;

;

;

;

;

;

;

;

;

;

; NUMBER OF SEQ ID NOS: 5

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

ALIGNMENTS

Query Match: 100.0%; Score: 12; DB 10; Length: 12;
Best Local Similarity: 100.0%; Pred. No. 1.2e-07;
Matches: 12; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;

QY 1 TSPINIHNGQKL 12
DB 1 TSPINIHNGQKL 12

RESULT 2
US-09-899-376-3
; Sequence 3, Application US/09899376
; Patent No. US20030102265A1
; GENERAL INFORMATION:
; APPLICANT: HONG, FRANK D.
; TITLE OF INVENTION: ISOLATION OF A CELL-SPECIFIC INTERNALIZING PEPTIDE THAT
; FILE REFERENCE: UTSC:645US

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00561

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117

SOFTWARE: animax Sequence Listing Engine vers. 1.1
SEQ ID NO 37407

LENGTH: 139
TYPE: PRT

ORGANISM: Homo sapiens
FEATURE:

OTHER INFORMATION: MAP TO AL022395.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.6

OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.9

OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.8

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.7

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.7

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.6

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.5

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.4

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5

OTHER INFORMATION: EXPRESSED IN TESTIS, SIGNAL = 5.00e-80

OTHER INFORMATION: SWISSPROT HIT: Q99593, EVALUE 7.20e-01

US-09-864-761-37407

Query Match 41.7%; Score 5; DB 10; Length 139;

Best Local Similarity 100.0%; Pred. No. 45; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 10
US-09-731-816-5

Query Match 41.7%; Score 5; DB 10; Length 139;
Best Local Similarity 100.0%; Pred. No. 45; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPLN 5
Db 13 TSPLN 17

RESULT 10
US-09-731-816-5

Query Match 41.7%; Score 5; DB 10; Length 139;
Best Local Similarity 100.0%; Pred. No. 45; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPLN 5
Db 13 TSPLN 17

RESULT 10
US-09-731-816-5

Query Match 41.7%; Score 5; DB 10; Length 139;
Best Local Similarity 100.0%; Pred. No. 45; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPLN 5
Db 13 TSPLN 17

RESULT 10
US-09-731-816-5

Query Match 41.7%; Score 5; DB 10; Length 139;
Best Local Similarity 100.0%; Pred. No. 45; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPLN 5
Db 13 TSPLN 17

RESULT 10
US-09-731-816-5

Query Match 41.7%; Score 5; DB 10; Length 139;
Best Local Similarity 100.0%; Pred. No. 45; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPLN 5
Db 13 TSPLN 17

US-09-731-816-5

Query Match 41.7%; Score 5; DB 10; Length 155;
Best Local Similarity 100.0%; Pred. No. 50; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LNIHN 8
Db 49 LNIHN 53

RESULT 11
US-09-886-404-5

Query Match 41.7%; Score 5; DB 10; Length 155;
Best Local Similarity 100.0%; Pred. No. 50; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LNIHN 8
Db 49 LNIHN 53

RESULT 11
US-09-886-404-5

Query Match 41.7%; Score 5; DB 10; Length 155;
Best Local Similarity 100.0%; Pred. No. 50; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LNIHN 8
Db 49 LNIHN 53

RESULT 11
US-09-886-404-5

Query Match 41.7%; Score 5; DB 10; Length 155;
Best Local Similarity 100.0%; Pred. No. 50; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LNIHN 8
Db 49 LNIHN 53

RESULT 11
US-09-886-404-5

Query Match 41.7%; Score 5; DB 10; Length 155;
Best Local Similarity 100.0%; Pred. No. 50; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LNIHN 8
Db 49 LNIHN 53

RESULT 11
US-09-886-404-5

Query Match 41.7%; Score 5; DB 10; Length 155;
Best Local Similarity 100.0%; Pred. No. 50; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LNIHN 8
Db 49 LNIHN 53

RESULT 11
US-09-886-404-5

Query Match 41.7%; Score 5; DB 10; Length 155;
Best Local Similarity 100.0%; Pred. No. 50; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LNIHN 8
Db 49 LNIHN 53

RESULT 11
US-09-886-404-5

Query Match 41.7%; Score 5; DB 10; Length 155;
Best Local Similarity 100.0%; Pred. No. 50; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LNIHN 8
Db 49 LNIHN 53

RESULT 11
US-09-886-404-5

Query Match 41.7%; Score 5; DB 10; Length 155;
Best Local Similarity 100.0%; Pred. No. 50; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LNIHN 8
Db 49 LNIHN 53

RESULT 11
US-09-886-404-5

Query Match 41.7%; Score 5; DB 10; Length 155;
Best Local Similarity 100.0%; Pred. No. 50; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LNIHN 8
Db 49 LNIHN 53

Fri Jan 3 14:42:47 2003

us-09-899-376-1.oli.rapb

Page 6

Qy	8	NGOKL	12
	1111		
Db	146	NGOKL	150

Search completed: January 3, 2003, 14:29:20
Job time: 10 secs

GenCore version 5.1.3
 Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 14:26:48 ; Search time 14 Seconds
 (without alignments)
 25.220 Million cell updates/sec

Title: US-09-899-376-1
 Perfect score: 12
 Sequence: 1 TSPINHNGOKL 12

Scoring table: OLIGO 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
 Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued-Patents-AA:*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep: *
 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep: *
 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep: *
 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep: *
 5: /cgn2_5/ptodata/1/iaa/PCUS_COMB.pep: *
 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep: *

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6 50.0	109	4	US-09-134-001C-3523	Sequence 3523, Appl
2	6 50.0	467	4	US-09-134-001C-3235	Sequence 3235, Appl
3	6 50.0	20	1	US-08-399-696-90	Sequence 90, Appl
4	5 41.7	59	1	US-08-399-696-101	Sequence 101, Appl
5	5 41.7	155	4	US-08-432-914A-8	Sequence 50, Appl
6	5 41.7	177	4	US-08-462-260A-50	Sequence 50, Appl
7	5 41.7	194	3	US-08-822-264-4	Sequence 4, Appl
8	5 41.7	194	4	US-09-565-808-21	Sequence 21, Appl
9	5 41.7	195	2	US-08-960-022-6	RESULT 2
10	5 41.7	195	4	US-09-565-808-2	US-09-134-001C-3235
11	5 41.7	220	3	US-08-822-264-1	Sequence 3235, Appl
12	5 41.7	223	3	US-08-822-264-3	Patent No. 6380370
13	5 41.7	223	4	US-09-565-808-22	GENERAL INFORMATION:
14	5 41.7	270	1	US-08-347-471-4	APPLICANT: Lynn Doucette-Stamm et al
15	5 41.7	277	1	US-08-347-471-2	TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
16	5 41.7	277	4	US-09-527-652A-26	TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
17	5 41.7	277	4	US-09-911-52A-27	FILE REFERENCE: GTC-007
18	5 41.7	277	4	US-09-911-52A-28	CURRENT APPLICATION NUMBER: US/09/134, 001C
19	5 41.7	278	1	US-08-414-926A-4	CURRENT FILING DATE: 1998-08-13
20	5 41.7	278	2	US-08-926-922-4	PRIOR APPLICATION NUMBER: US 60/064, 964
21	5 41.7	278	3	US-09-253-682-4	PRIOR FILING DATE: 1997-11-08
22	5 41.7	278	4	US-09-527-652A-25	PRIOR APPLICATION NUMBER: US 60/055, 779
23	5 41.7	285	4	US-09-911-52A-25	PRIOR FILING DATE: 1997-08-14
24	5 41.7	301	2	US-08-783-464A-1	NUMBER OF SEQ ID NOS: 5674
25	5 41.7	349	3	US-08-603-150A-17	SEQ ID NO 3235
26	5 41.7	354	4	US-09-067-089-2	LENGTH: 467
27	5 41.7	356	1	US-08-270-583-2	TYPE: PRT

ALIGNMENTS

; ORGANISM: *Staphylococcus epidermidis*
; US-09-134-001C-3235

APPLICATION NUMBER: US/08/399, 696
FILING DATE: 02-MAR-1995

```

Query Match          50.0%;  Score 6;  DB 4;  Length 467;
Best Local Similarity 100.0%;  Pred. No. 26;
Matches 6;  Conservative 0;  Mismatches 0;  Indels 0;
Qy      5 NIHNGQ 10
       ||||| |
Db      456 NIHNGQ 461

```

PATENT APPLICATION DATA:
APPLICATION NUMBER: US 08/156,671
FILING DATE: 22-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Smith, William M
REGISTRATION NUMBER: 30,223
REFERENCE/DOCKET NUMBER: 15522-000710
TELECOMMUNICATION INFORMATION:

FILING DATE: 02-MAY-1995
 CLASSIFICATION: 424
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/450,846
 FILING DATE: 27-MAY-1994
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/177,747
 FILING DATE: 14-JUN-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: CHING, EDWIN P.
 REGISTRATION NUMBER: 34,090
 REFERENCE/DOCKET NUMBER: DX0388K3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-52-9196
 TELEFAX: 415-496-1200
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 155 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-432-994A-8

Query Match 41.7%; Score 5; DB 4; Length 155;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 GENERAL INFORMATION:
 APPLICANT: JOHN N. SIMONS
 APPLICANT: TAMI J. PILOT-MATIAS
 APPLICANT: GEORGE J. DAWSON
 APPLICANT: GEORGE G. SCHLAUDER
 APPLICANT: SURESH M. DESAI
 APPLICANT: THOMAS P. LEARY
 APPLICANT: ANTHONY SCOTT MUEHRUFF
 APPLICANT: JAMES C. ERKER
 APPLICANT: SHERI L. BUIJK
 APPLICANT: ISA K. MUSHAHWAR
 TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
 NUMBER OF SEQUENCES: 716
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: ABBOTT LABORATORIES
 STREET: 100 ABBOTT PARK ROAD
 CITY: ABBOTT PARK
 STATE: IL
 COUNTRY: USA
 ZIP: 60064-3500

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PATENTIN Release #1.0, version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/469,260A
 FILING DATE: 05-JAN-1994
 CLASSIFICATION:
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US/08/424,550
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:

NAME: POREMBSKI, PRISCILLA E.
 REGISTRATION NUMBER: 33,207
 REFERENCE/DOCKET NUMBER: 5527.PC.01
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 708-938-5623
 TELEFAX: 708-938-5623
 INFORMATION FOR SEQ ID NO: 50:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 177 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-469-260A-50

Query Match 41.7%; Score 5; DB 4; Length 177;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 GENERAL INFORMATION:
 APPLICANT: Hilmann, Jennifer L.
 APPLICANT: Murry, Lynn E.
 TITLE OF INVENTION: NOVEL HUMAN CYTOKINE/STEROID
 TITLE OF INVENTION: RECEPTOR PROTEIN
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: US
 ZIP: 94304

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASSEQ Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/822,264
 FILING DATE: 05-JAN-1994
 CLASSIFICATION: 530
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0233 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-845-4166
 TELEX:
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 194 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: GenBank
 CLONE: 1657409
 US-08-822-264-4

Query Match 41.7%; score 5; DB 3; Length 194;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0;
 Gaps 0;

Qy 1 TSPLN 5
 Db 27 TSPLN 31

RESULT 8
 US-09-565-808-21
 Sequence 21, Application US/09565808
 ; GENERAL INFORMATION:
 ; PATENT NO.: 6432674
 ; APPLICANT: Hirata, Yuichi
 ; TITLE OF INVENTION: STEROID HORMONE BINDING PROTEIN
 ; FILE REFERENCE: 06501-059001 WO/JP98/05010
 ; CURRENT APPLICATION NUMBER: US/09/565, 808
 ; CURRENT FILING DATE: 2000-05-05
 ; PRIORITY NUMBER: WO/JP98/05010
 ; PRIORITY FILING DATE: 1998-11-06
 ; PRIORITY APPLICATION NUMBER: JP/9/322376
 ; PRIORITY FILING DATE: 1997-11-07
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO: 21
 ; LENGTH: 194
 ; TYPE: PRT
 ; ORGANISM: Suidae

Query Match 41.7%; score 5; DB 4; Length 194;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0;
 Gaps 0;

Qy 1 TSPLN 5
 Db 27 TSPLN 31

RESULT 9
 US-09-960-022-6
 Sequence 6, Application US/08960022
 ; GENERAL INFORMATION:
 ; PATENT NO.: 5976837
 ; APPLICANT: Jacobs, Kenneth
 ; APPLICANT: McCoy, John M.
 ; APPLICANT: Lavalle, Edward R.
 ; APPLICANT: Racine, Lisa A.
 ; APPLICANT: Merberg, David
 ; APPLICANT: Tracy, Maurice
 ; APPLICANT: Spaulding, Vicki
 ; APPLICANT: Agostino, Michael J.
 ; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
 ; NUMBER OF SEQUENCES: 30
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genetics Institute, Inc.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: MA
 ; COUNTRY: U.S.A.
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; COMPUTER TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/960,022
 ; FILING DATE:
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:

Query Match 41.7%; score 5; DB 2; Length 195;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0;
 Gaps 0;

Qy 1 TSPLN 5
 Db 27 TSPLN 31

RESULT 10
 US-09-565-808-2
 Sequence 2, Application US/09565808
 ; GENERAL INFORMATION:
 ; PATENT NO.: 6432674
 ; APPLICANT: Hirata, Yuichi
 ; TITLE OF INVENTION: STEROID HORMONE BINDING PROTEIN
 ; FILE REFERENCE: 06501-059001 WO/JP98/05010
 ; CURRENT APPLICATION NUMBER: US/09/565, 808
 ; CURRENT FILING DATE: 2000-05-05
 ; PRIORITY NUMBER: WO/JP98/05010
 ; PRIORITY FILING DATE: 1998-11-06
 ; PRIORITY APPLICATION NUMBER: JP/9/322376
 ; PRIORITY FILING DATE: 1997-11-07
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO: 2
 ; LENGTH: 195
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens

Query Match 41.7%; score 5; DB 4; Length 195;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0;
 Gaps 0;

Qy 1 TSPLN 5
 Db 27 TSPLN 31

RESULT 11
 US-09-822-564-1
 Sequence 1, Application US/09822264
 ; GENERAL INFORMATION:
 ; APPLICANT: Gol, Surya K.
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Murry, Lynn E.
 ; TITLE OF INVENTION: NOVEL HUMAN CYKOKINE/STEROID
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: US
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/347,471
 FILING DATE: 08-DECEMBER-94
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: NO. 5710015th, Robert J.
 REGISTRATION NUMBER: 27,366
 REFERENCE/DOCKET NUMBER: T-1160
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (908) 594-7262
 TELEX: (908) 594-4720
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 270 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

Query Match 41.7%; Score 5; DB 1; length 270;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 5; Conservative 0; Mismatches 0;
 Mismatches 5; Conservative 0; Mismatches 0;
 Indels 0; Gaps 0;
 Qy 8 NGOKL 12
 |||||
 Db 142 NGOKL 146

RESULT 15
 US-08-347-471-2
 Sequence 2, Application US/08347471
 Patent No. 5710015
 GENERAL INFORMATION:

APPLICANT: McAllister, George
 TITLE OF INVENTION: cDNA CLONING OF INOSITOL MONOPHOSPHATASE
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Robert J. NO. 5710015th
 STREET: P.O. Box 2000, 126 E. Lincoln Avenue
 CITY: Railway
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07005-0907
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/347,471
 FILING DATE: 08-DECEMBER-94
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: NO. 5710015th, Robert J.
 REGISTRATION NUMBER: 27,366
 REFERENCE/DOCKET NUMBER: T-1160
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (908) 594-7262
 TELEX: (908) 594-4720
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 277 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

Query Match 41.7%; Score 5; DB 1; length 277;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 5; Conservative 0; Mismatches 0;
 Mismatches 5; Conservative 0; Mismatches 0;
 Indels 0; Gaps 0;
 Qy 8 NGOKL 12
 |||||
 Db 142 NGOKL 146

Search completed: January 3, 2003, 14:29:03
 Job time : 15 secs

	Matches	Conservative	Mismatches	Indels	Gaps	0;
Qy	8	NGOKL 12				
Db	142	NGOKL 146				

Copyright (c) 1993 - 2003 Compugen Ltd.

Gencore version 5.1.3

Run on: January 3, 2003, 14:26:13 ; Search time 28 Seconds
(without alignments)

88.306 Million cell updates/sec

OM protein - protein search, using sw mode

Title: US-09-899-376-1

Perfect score: 12

Sequence: 1 TSPLNTHNGQKL 12

Scoring table: ORIGO

Word size : 0

gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_21; *

1: sp_archea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rabbit:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteria:*

17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query	Match Length	DB ID	Description
1	7	58.3	173	4 O9UKC5	O9UKC5 homo sapien
2	7	58.3	188	2 O93MT1	O93MT1 photorhabdus
3	7	58.3	321	4 O96ER2	O96ER2 homo sapien
4	7	58.3	367	11 O91V9	O91V9 mus musculus
5	7	58.3	415	11 O9JIE4	O9JIE4 mus musculus
6	7	58.3	471	4 O9HOV2	O9HOV2 homo sapien
7	7	58.3	480	11 O92J5	O92J5 mus musculus
8	7	58.3	480	11 O92J5	O92J5 mus musculus
9	6	50.0	158	10 O8VX3	O8VX3 arabidopsis
10	6	50.0	169	2 O51528	O51528 pseudomonas
11	6	50.0	169	16 O9HNU0	O9HNU0 pseudomonas
12	6	50.0	393	2 O9JRQ0	O9JRQ0 actinobacillus
13	6	50.0	420	50.0 O17057	O17057 caenorhabditis
14	6	50.0	706	12 O89581	O89581 african swi
15	6	50.0	706	12 O89581	O89581 african swi
16	6	50.0	1050	3 O00666	O00666 emericella

ALIGNMENTS

RESULT 1

ID	Q9UKC5	PRELIMINARY:	PRT:	173 AA.
AC	Q9UKC5;			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	F-box protein Fbx3 (Fragment).			
GN	Fbx3.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TAXID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE:2000060; PubMed=10331035;			
RA	Cenciarelli C., Chiaru D.S., Guardavaccaro D., Parks W., Vidal M.,			
RA	Ragno M.;			
RT	"Identification of a family of human F-box proteins."			
RL	Curr. Biol. 9:1177-1179(1999).			
DR	AF174555; AF04516.1; -.			
DR	InterPro; IPR01810; F-box.			
DR	Pfam; PF00646; F-box; 1.			
DR	SMART; SM00256; FBOX; 1.			
DR	PROSITE; P550181; FBOX; 1.			
FT	NON_TER 1 1			
SEQUENCE	173 AA; 20087 MW; DA01B222A57DAA4 CRC64;			

Query Match 58.3%; Score 7; DB 4; Length 173; Best local Similarity 100.0%; Pred. No. 1.2; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 IHNQKL 12

Db 146 IHNQKL 152

RESULT 2

Q3MTL	
1	111111 -

Q91V9	PRELIMINARY;	PRT;	367 AA.
Q91V9;			
AC			
01-DEC-2001 (TREMBLrel. 19, Created)			
DT			
01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT			
01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE			
MifF			
GN			
MRFF.			
Photorhabdus luminescens (Xenorhabdus luminescens).			
OS			
Photorhabdus			
OC			
Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC			
NCBI_TAXID=29488;			
RN			
[1]			
SEQUENCE FROM N.A.			
RP			
STRAIN=K12;			
RC			
RA			
Mélet L., Pimenta A., Blight M.:			
RT			
"Cloning and analysis of a mannose resistant fimbrial operon from the			
entomopathogenic bacterium, Photorhabdus luminescens K12.;"			
RL			
Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.			
DR			
EMBL: AP39603; AAK8248.1; -.			
DR			
InterPro; IPR00219; Fimbrial.			
DR			
Pfam: PF00419; Fimbrial; 1.			
DR			
PRINTS: PRO1613; FIMBRIALPAGE.			
SEQUENCE 188 AA; 20383 MW; CE6DE889285C40CB CRC64;			
SQ			
Query Match 58.3%; Score 7; DB 2; Length 188;			
Best Local Similarity 100.0%; Pred. No. 1.3;			
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY			
6 IHNGQKL 12			
Db			
138 IHNGQKL 144			
RESULT 3			
Q95FR2	PRELIMINARY;	PRT;	321 AA.
ID			
Q95FR2			
AC			
Q95FR2;			
DT			
01-DEC-2001 (TREMBLrel. 19, Created)			
DT			
01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT			
01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE			
DE			
Homo sapiens (Human).			
OS			
OC			
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Buteleostomi;			
OC			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
NCBI_TAXID=9606;			
OX			
RN			
[1]			
SEQUENCE FROM N.A.			
RC			
SEQUENCE FROM N.A.			
RA			
Strausberg R.:			
RL			
Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.			
DR			
EMBL; BCO12012; AAH12012.1; -.			
DR			
InterPro; IPR001909; KRAB.			
DR			
InterPro; IPR000322; Znf_C2H2.			
DR			
Pfam; PF01352; KRAB; 1.			
DR			
Pfam; PR0096; zf-C2H2; 7.			
DR			
ProDom; PD00003; Znf_C2H2; 6.			
DR			
PROSITE; PS58085; KRAB; 1.			
DR			
PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_6.			
DR			
PROSITE; PS51057; ZINC_FINGER_C2H2_2; 7.			
KW			
DNA-binding; Zinc-finger.			
SEQUENCE 321 AA; 36369 MW; 62BE342C8B7067D9 CRC64;			
SQ			
Query Match 58.3%; Score 7; DB 4; Length 321;			
Best Local Similarity 100.0%; Pred. No. 1.9;			
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY			
6 IHNGQKL 12			
Db			
313 IHNGQKL 319			
RESULT 4			
Q91V9			
RESULT 6			

09H0V2	PRELIMINARY;	PRT;	471 AA.
ID 09H0V2			
AC 09H0V2;			
DT 01-MAR-2001 (TREMBLrel. 16, Created)			
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE Hypothetical 54.6 kDa protein.			
GN DKFZP54B092.			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarhini; Homidae; Homo.			
RA NCBI_TAXID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC TISSUE=BRAIN;			
RX MEDLINE=21154917; PubMed=11230166;			
RA Wielmann S., Weil B., Wellenreuther R., Gasssenhuber J., Glassl S., Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N., Ansorge W., Boecker M., Bloecker H., Bauersachs S., Blum H., Mewes H.W., Ottenwalder B., Obermaier B., Tampe J., Heubner D., Wambutt R., Korn B., Klein M., Pousta R.			
RT "Towards a Catalog of Human Genes and Proteins: Sequencing and Analysis of 500 Novel Complete Protein Coding Human cDNAs.;"			
RL Genome Res. 11:42-43(2001).			
DR EMBL: AL136525; CAB6560.1; -.			
DR InterPro; IPR001810; F-box.			
DR Pfam; PF00546; F-box; 1.			
DR SMART; SM00256; FBOX; 1.			
DR PROSITE; PS50181; FBOX; 1.			
KW Hypothetical protein.			
SO SEQUENCE 471 AA; 54572 MN; CCFD0DDC8A884410 CRC64;			
Query Match 58.3%; Score 7; DB 4; Length 471; Best Local Similarity 100.0%; Pred. No. 2, 6; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY 6 IHNGQKL 12			
Db 148 IHNGQKL 154			
RESULT 7			
ID 09DC63	PRELIMINARY;	PRT;	480 AA.
AC 09DC63;			
DT 01-JUN-2001 (TREMBLrel. 17, Created)			
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE 1200002609Rik protein.			
GN FBXO3 OR 1200002609Rik.			
OS Mus musculus (Mouse).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
RA NCBI_TAXID=10090;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=C57BL/6J; TISSUE=TESTIS;			
RX MEDLINE=21085660; PubMed=11217851;			
RA Kawaji J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Iwasa M., Nishi K., Kyosawa H., Kondo S., Yamamoto I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batyalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Rono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Ronald M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Garibaldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombberts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sakai H., Sato K., Schoenbach Y., Seva T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.-H., Weitz C., Whittaker C., Wilming L., Wynnshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S., Hayashizaki Y.			
RA "Functional annotation of a full-length mouse cDNA collection.;"			
RT Nature 409:685-690(2000).			
RL MEDP; MGII:1929084; Fbxo3.			
DR InterPro; IPR001810; F-box.			
DR Pfam; PF00546; F-box; 1.			
DR SMART; SM00256; FBOX; 1.			
DR PROSITE; PS50181; FBOX; 1.			
SO SEQUENCE 480 AA; 55227 MN; 9AD95505076A67787 CRC64;			
Query Match 58.3%; Score 7; DB 11; Length 480; Best Local Similarity 100.0%; Pred. No. 2, 6; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY 6 IHNGQKL 12			

Db	148 IHNQQL 154	Qy	4 LNHNG 9
RESULT 9			
Q8VXX3	PRELIMINARY;	PRT;	158 AA.
AC	08VXX3;		
DT	01-MAR-2002 (TREMBLrel. 20, Created)	ID	Q9HVN0
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)	ID	Q9HVN0;
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)	AC	Q9HVN0;
DE	Hypothetical 18.2 kDa protein.	DT	01-MAR-2001 (TREMBLrel. 16, Created)
GN	AT5G58740;	DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)
OS	Arabidopsis thaliana (Mouse-ear cress).	DE	01-MAR-2002 (TREMBLrel. 20, Last annotation update)
OC	Eukaryota; Viridiplantida; Streptophyta;	TYPE	4 Fimbrial biogenesis protein FlmR.
OC	Spermatophyta; Magnoliophyta; eudicots; Rosidae;	GN	PFM7 OR PA5749.
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.	OS	Pseudomonas aeruginosa.
OX	NCBI_TAXID=3702;	OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
RN	[1]	OC	Pseudomonas.
RP	SEQUENCE FROM N.A.	OX	NCBI_TAXID=287;
RA	Yanada K., Bahn J., Chan M.M., Chang C.H., Chang E., Dale J.M.,	RN	[1]
RA	Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,	RP	SEQUENCE FROM N.A.
RA	Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,	RC	STRAIN=ATCC 15692 / PA01;
RA	Kaminci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,	RX	MEDLINE:20437337; Published=10984043;
RA	Kamiya A., Karlin-Neumann G., Kawai J., Lin J.,	RA	Stover C.K., Puhar X.-P., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA	Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,	RA	Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA	Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,	RA	Garber R.L., Gooley L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA	Davis R.W., Ecker J.R., Theologis A.,	RA	Bordy L.L., Coulter S.N., Folgar K.R., Kas A., Larbig K., Lim R.M.,
RT	"Full length cDNA of gene At5g58740 (GI:15237742)." ;	RA	Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RL	Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.	RA	Reizer J., Salter M.H., Hancock R.E.W., Lory S., Olson M.V.,
EMBL	AY074384; AAL67080.1; -.	RT	"Complete genome sequence of Pseudomonas aeruginosa PA01, an
DR	InterPro; IPR005307; Nuclear_move; 1.	RT	opportunistic pathogen." ;
DR	PF03593; Nuclear_move; 1.	DR	Nature 406:959-964 (2000).
KW	Hypothetical protein.	DR	EMBL; AB004869; AAG0737.1; -.
SQ	SEQUENCE 158 AA; 18211 MW; 8764B24FCC8F1BF3 CRC64;	Qy	SEQUENCE 169 AA; 18889 MW; EB08764FAFF2AAEC CRC64;
Query	Match	50.0%	50.0%
Best	Local Similarity	100.0%	Score 6; DB 10; Length 158;
Matches	6; Conservative	0;	Pred. No. 16; Mismatches 0; Indels 0; Gaps 0;
Oy	6 IHNQQL 11	Db	123 LNHNG 128
Db	[14] IHNQQL 19		
RESULT 10			
Q51528	PRELIMINARY;	PRT;	169 AA.
ID	Q51528;		
AC	051528;		
DT	01-NOV-1996 (TREMBLrel. 01, Created)	ID	Q9TRQ0
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)	ID	Q9TRQ0;
DT	01-NOV-1998 (TREMBLrel. 08, Last annotation update)	AC	Q9TRQ0;
DE	Fimbrial protein FIMT.	DT	01-OCT-2000 (TREMBLrel. 15, Created)
GN		DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)
OS	Pseudomonas aeruginosa.	DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)
OS	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;	DE	Putative glycosyltransferase.
OC	Pseudomonas.	OS	Actinobacillus actinomycetemcomitans (Haemophilus
OC	OC	actinomycetemcomitans).	
OX	OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;	
RN	OC	Actinobacillus.	
RP	SEQUENCE FROM N.A.	OX	NCBI_TAXID=714;
RC	STRAIN=PA01;	RN	[1]
RX	MEDLINE:6227255; PubMed=8682785;	RP	SEQUENCE FROM N.A.
RA	Alm R.A., Mattick J.S.;	RC	STRAIN=CU1000;
RT	"Identification of two genes with preplin-like leader sequences involved in type 4 fimbrial biogenesis in Pseudomonas aeruginosa." ;	RX	MEDLINE:21391794; PubMed=11500407;
RT	J. Bacteriol. 178:3805-3817 (1996).	RA	Kaplan J.B., Perry M.B., MacLean L.L., Furgang D., Wilson M.E.,
RL	FUNCTION: INVOLVED IN TYPE 4 FIMBRIAL BIOGENESIS.	RA	Fine D.H.,
CC		RT	"Structural and Genetic Analyses of O Polysaccharide from
DR	EMBL; L48934; AAB39270.1; -.	RT	Actinobacillus actinomycetemcomitans Serotype f." ;
KW	Fimbria.	RL	Infect. Immun. 69:5375-5384 (2001).
SQ	SEQUENCE 169 AA; 18958 MW; EB08764E1FF2AAEC CRC64;	DR	EMBL; AR213680; AAF25887.2; -.
Query	Match	50.0%	50.0%
Best	Local Similarity	100.0%	Score 6; DB 2; Length 169;
Matches	6; Conservative	0;	Pred. No. 17; Mismatches 0; Indels 0; Gaps 0;

Matches	6	Conservative	0	Mismatches	0	Indels	0	Gaps	0
OY	4	LNIHNG	9						
ID									
Db	221	LNIHNG	226						
RESULT	13								
O17057		PRELIMINARY;		PRT;	420 AA.				
ID	017057								
AC	017057;								
DT	01-JUN-1998	(T-EMBLel. 05, Created)							
DT	01-OCT-2001	(T-EMBLel. 18, Last sequence update)							
DT	01-DEC-2001	(T-EMBLel. 19, Last annotation update)							
DE		Hypothetical	46.4 kDa protein.						
GN	C24A1.2.								
OS		Caenorhabditis elegans.							
OC		Bukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Pejoderrinae; Caenorhabditis.							
OX		Rhabditidae							
OX		NBGI-TaxID=6239;							
RN	{1}								
RP		SEQUENCE FROM N.A.							
RC		STRAIN=BRISTOL N2;							
RA		None;							
RT		"Genome sequence of the nematode <i>C. elegans</i> : a platform for investigating biology. The <i>C. elegans</i> Sequencing Consortium.";							
RL		Science 283:2012-2018(1998).							
RN	[2]								
RP		SEQUENCE FROM N.A.							
RC		STRAIN=BRISTOL N2;							
RA		Connell M.;							
RT		"The sequence of <i>C. elegans</i> cosmid C24A1.1";							
RL		Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.							
RN	[3]								
RP		SEQUENCE FROM N.A.							
RC		STRAIN=BRISTOL N2;							
RA		Waterson R.;							
RT		"Direct Submission.";							
RL		Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.							
DR		EMBL: AF04491; ARB0311.2; -.							
DR		HSSP: PI4921; 25TT.							
DR		InterPro: IPR002341; ETS.							
DR		InterPro: IPR002341; HSF-ETS.							
DR		PFam: PF00178; Ets; 1.							
DR		PRINTS: PR00454; ETSDOMAIN.							
DR		SMART: SM00413; ETS; 1.							
DR		PROSITE: PS00345; ETS_DOMAIN_1; 1.							
DR		PROSITE: PS00346; ETS_DOMAIN_2; 1.							
DR		PROSITE: PS00561; ETS_DOMAIN_3; 1.							
KW		Hypothetical protein.							
SO		SEQUENCE 420 AA; 46383 MW: AA1DF31254563459 CRC64;							
Query		Match	50.0%	Score	6	DB	5	Length	420;
Best		Local	Similarity	100.0%	Pred.	No.	33;	Mismatches	
Matches	6	Conservative	0		0;	Indels	0	Gaps	0;
OY	1	TSPLNI	6						
Db	21	TSPLNI	26						
RESULT	14								
O89581		PRELIMINARY;		PRT;	705 AA.				
ID	089581								
AC	089581;								
DT	01-NOV-1996	(T-EMBLel. 01, Created)							
DT	01-NOV-1995	(T-EMBLel. 01, Last sequence update)							
DT	01-DEC-2001	(T-EMBLel. 19, Last annotation update)							
DE		HELICASE.							
GN	0705L								
OS		African swine fever virus (ASFV).							
Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.									
OX		NCBI_TAXID=10497;							
RN	[1]								
RP		SEQUENCE FROM N.A.							
RC		STRAIN=BA71V;							
RX		MEDLINE=96036500; PubMed=7483270;							
RA		Yanez R. J., Rodriguez J. M., Nogal M. L., Yuste L., Enriquez C., Rodriguez J. F., Vinuela E.;							
RA		"Immune protection conferred by the baculovirus-related glycoprotein of Theogoto virus (Orthomyxoviridae).";							
RT		Virology 208:249-278(1995).							
RN	[2]								
RP		SEQUENCE FROM N.A.							
RC		STRAIN=BA71V;							
RX		MEDLINE=9423765; PubMed=8178480;							
RA		La Vega I., Gonzalez A., Blasco R., Calvo V., Vinuela E.;							
RT		"Nucleotide sequence and variability of the inverted terminal repetitions of African swine fever virus DNA.,"							
RL		Virology 201:152-156(1994).							
RN	[3]								
RP		SEQUENCE FROM N.A.							
RC		STRAIN=BA71V;							
RX		MEDLINE=90219205; PubMed=2325203;							
RA		Gonzalez A., Calvo V., Almazan F., Almendral J. M., Ramirez J. C., La Vega I., Blasco R., Vinuela E.;							
RT		"Multigene families in African swine fever virus: family 360.,"; J. Virol. 64:2073-2081(1990).							
RN	[4]								
RP		SEQUENCE FROM N.A.							
RC		STRAIN=BA71V;							
RX		MEDLINE=90219204; PubMed=2325202;							
RA		Almendral J. M., Almazan F., Blasco R., Vinuela E.;							
RT		"Multigene families in African swine fever virus: family 110.,"; J. Virol. 64:2064-2072(1990).							
RN	[5]								
RP		SEQUENCE FROM N.A.							
RC		STRAIN=BA71V;							
RX		MEDLINE=9113498B; PubMed=1994575;							
RA		Camacho A., Vinuela E.;							
RT		"Protein p22 of African swine fever virus: an early structural protein that is incorporated into the membrane of infected cells.,"; Virology 181:251-257(1991).							
RN	[6]								
RP		SEQUENCE FROM N.A.							
RC		STRAIN=BA71V;							
RX		MEDLINE=94187118; PubMed=8139051;							
RA		Almazan F., Murquia J. R., Rodriguez J. M., Vinuela E.;							
RA		Rodríguez J. M., Yanez R. J., Pan R., Rodriguez J. F., Salas M. L., Vinuela E.;							
RA		"Multigene families in African swine fever virus: family 505.,"; J. Virol. 68:2740-2751(1994).							
RN	[8]								
RP		SEQUENCE FROM N.A.							
RC		STRAIN=BA71V;							
RX		MEDLINE=94187118; PubMed=8139051;							
RA		Yanez R. J., Rodriguez J. M., Yanez R. J., Vinuela E.;							
RT		"African swine fever virus thymidine kinase gene: sequence and transcriptional mapping.,"; J. Gen. Virol. 74:1633-1638(1993).							
RL		[9]							
RN		SEQUENCE FROM N.A.							
RC		STRAIN=BA71V;							
RX		MEDLINE=94055656; PubMed=8245848;							
RA		Alcami A., Angulo A., Vinuela E.;							
RT		"Mapping and sequence of the gene encoding the African swine fever virus protein of M(r) 11000.,"; J. Gen. Virol. 74:2317-2324(1993).							
RL		[10]							
RP		SEQUENCE FROM N.A.							
RC		STRAIN=BA71V;							

RX MEDLINE=93277388; PubMed=8503790;
 RA Munoz M., Freije J.M., Vinuela E., Lopez-Otin C.;
 RT "Structure and expression in *E. coli* of the gene coding for protein
 p10 of African swine fever virus.";
 RL Arch. Virol. 130:93-107(1993).
 RT [11]
 RP SEQUENCE FROM N A.
 RC STRAIN=BAT1V;
 RX MEDLINE=90357780; PubMed=2389555;
 RA Blasco R., Lopez-Otin C., Munoz M., Bockamp E.O., Simon-Mateo C.,
 Vinuela E.;
 RT "Sequence and evolutionary relationships of African swine fever virus
 thymidine kinase.";
 RL Virology 178:301-304(1990).
 RN [12]
 RP SEQUENCE FROM N A.
 RC STRAIN=BAT1V;
 RX MEDLINE=93381390; PubMed=8506138;
 RA Yanez R.J., Boursnell M., Nogal M.L., Vistre L., Vinuela E.;
 RT "African swine fever virus encodes two genes which share significant
 homology with the two largest subunits of DNA-dependent RNA
 polymerases.;"
 RL Nucleic Acids Res. 21:2423-2427(1993).
 RN [13]
 RP SEQUENCE FROM N A.
 RC STRAIN=BAT1V;
 RX MEDLINE=93353606; PubMed=8102411;
 RA Rodriguez J.M., Yanez R.J., Almazan F., Vinuela E., Rodriguez J.F.;
 RT "African swine fever virus encodes a CD2 homolog responsible for the
 adhesion of erythrocytes to infected cells.";
 RL J. Virol. 67:5312-5320(1993).
 RN [14]
 RP SEQUENCE FROM N A.
 RC STRAIN=BAT1V;
 RX MEDLINE=91085774; PubMed=8262374;
 RA Yanez R.J., Rodriguez J.M., Boursnell M., Rodriguez J.F., Vinuela E.;
 RT "Two putative African swine fever virus helicases similar to yeast
 DEAH, pre-mRNA processing proteins and vaccinia virus ATPases DIL
 and DCR.";
 RL Gene 134:161-174(1993).
 RN [15]
 RP SEQUENCE FROM N A.
 RC STRAIN=BAT1V;
 RX MEDLINE=90223993; PubMed=2327074;
 RA Lopez-Otin C., Freije J.M., Parra F., Mendez E., Vinuela E.;
 RT "Mapping and sequence of the gene coding for protein p72, the major
 capsid protein of African swine fever virus.";
 RL Virology 175:477-484(1990).
 RN [16]
 RP SEQUENCE FROM N A.
 RC STRAIN=BAT1V;
 RX MEDLINE=94123986; PubMed=8293992;
 RA Rodriguez J.M., Yanez R.J., Rodriguez J.F., Vinuela E., Salas M.L.;
 RT "The DNA polymerase-encoding gene of African swine fever virus:
 sequence and transcriptional analysis.";
 RL Gene 136:103-110(1993).
 RN [17]
 RP SEQUENCE FROM N A.
 RC STRAIN=BAT1V;
 RX MEDLINE=93327788; PubMed=8335009;
 RA Simon-Mateo C., Andres G., Vinuela E.;
 RT "Polyprotein processing in African swine fever virus: a novel gene
 expression strategy for a DNA virus.";
 RL EMBO J. 12:2977-2987(1993).
 RN [18]
 RP SEQUENCE FROM N A.
 RC STRAIN=BAT1V;
 RX MEDLINE=9333210; PubMed=8474154;
 RA Prados F.J., Vinuela E., Alcami A.;
 RT "Sequence and characterization of the major early phosphoprotein p32
 of African swine fever virus.";
 RL J. Virol. 67:2475-2485(1993).
 RN [19]

RX SEQUENCE FROM N A.
 RC STRAIN=BAT1V;
 RX MEDLINE=92260660; PubMed=1583732;
 RA Alcami A., Angulo A., Lopez-Otin C., Munoz M., Freije J.M.,
 Carrascosa A.L., Vinuela E.;
 RA "Amino acid sequence and structural properties of protein p12, an
 African swine fever virus attachment protein.";
 RT J. Virol. 66:3860-3868(1992).
 RN [20]
 RP SEQUENCE FROM N A.
 RC STRAIN=BAT1V;
 RX MEDLINE=93174975; PubMed=8438592;
 RA Yanez R.J., Vinuela E.;
 RT "African swine fever virus encodes a DNA ligase.";
 RL Virology 193:531-536(1993).
 RN [21]
 RP SEQUENCE FROM N A.
 RC STRAIN=BAT1V;
 RX MEDLINE=93174941; PubMed=8382999;
 RA pena L., Yanez R.J., Revilla Y., Vinuela E., Salas M.L.;
 RT "African swine fever virus guanylyltransferase.";
 RL Virology 193:319-328(1993).
 RN [22]
 RP SEQUENCE FROM N A.
 RC STRAIN=BAT1V;
 RX MEDLINE=95159428; PubMed=7856088;
 RA Simon-Mateo C., Freije J.M., Andres G., Lopez-Otin C., Vinuela E.;
 RT "Mapping and sequence of the gene encoding protein p17, a major
 African swine fever virus structural protein.";
 RL Virology 206:1140-1144(1995).
 RN [23]
 RP SEQUENCE FROM N A.
 RC STRAIN=BAT1V;
 RX MEDLINE=92263807; PubMed=1316688;
 RA Garcia-Beato R., Freije J.M., Lopez-Otin C., Blasco R., Vinuela E.,
 Salas M.L.;
 RT "A gene homologous to topoisomerase II in African swine fever virus.";
 RL Virology 188:935-947(1992).
 RN [24]
 RP SEQUENCE FROM N A.
 RC STRAIN=BAT1V;
 RX MEDLINE=94091056; PubMed=8266720;
 RA Freije J.M., Lain S., Vinuela E., Lopez-Otin C.;
 RT "Nucleotide sequence of a nucleoside triphosphate phosphohydrolase
 gene from African swine fever virus.";
 RL Query Match 50.0%; score 6; DB 12; Length 706;
 RT Best Local Similarity 100.0%; Fred. No. 49;
 RT Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 SPLNTH 7
 DB 485 SPLNTH 490

RESULT 15
 Q89926 ID 089925 PRELIMINARY; PRT; 706 AA.
 AC 089925
 DT 01-NOV-1995 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HELICASE.
 GN J10L.
 OS African swine fever virus (ASFV).
 OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfarvirus.
 OX NCBI TaxID=10497;
 RN [1]
 RP SEQUENCE FROM N A.
 RC STRAIN=MALAWI_L1L20_1/
 RX MEDLINE=9404995; PubMed=8409937;
 RA Vydelingum S., Baylis S.A., Bristow C., Smith G.L., Dixon L.K.;
 RT "Duplicated genes within the variable right end of the genome of a

RT pathogenic isolate of African swine fever virus.";
 J. Gen. Virol. 74:2125-2130(1993).
 RL
 RN [2]
 RP
 RC
 STRAIN=MALAWI_LIL20 /1;
 RA
 RA
 Dixon L.K., Twigg S.R.F., Baylis S.A., Vydelingum S., Bristow C.,
 Hammond J.M., Smith G.L.;
 RT "Nucleotide sequence of a 55 kbp region from the right end of the
 genome of a pathogenic African swine fever virus isolate (Malawi
 LIL20/1).";
 RT
 RL
 RN
 RP
 SEQUENCE FROM N.A.
 RX MEDLINE=93393435; PubMed=8397501;
 RA Roberts P.C., Lu Z., Kutish G.F., Rock D.L.;
 RT "Three adjacent genes of African swine fever virus with similarity to
 essential poxvirus genes.";
 RL Arch. Virol. 132:331-342(1993).
 RN
 RP
 RC
 STRAIN=MALAWI;
 RX MEDLINE=9339456; PubMed=8376971;
 RA Baylis S.A., Twigg S.R.F., Vydelingum S., Dixon L.K., Smith G.L.;
 RT "Three African swine fever virus genes encoding proteins with homology
 to putative helicases of vaccinia virus.";
 RL J. Gen. Virol. 74:1969-1974(1993).
 DR EMBL; X71182; CAA50820.1; -.
 DR EMBL; M88275; AAA03222.1; -.
 DR EMBL; X72952; CAA51457.1; -.
 DR InterPro; IPR01650; Helicase_C.
 DR Pfam; PF00271; helicase_C; 1.
 DR SMART; SM00490; HELICC; 1.
 KW ATP binding; Helicase.
 SQ SEQUENCE 706 AA; 80498 MW; 0D22C089781FAA46 CRC64;

Query Match 50.0%; Score 6; DB 12; Length 706;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 SPINIH 7
 Db 485 SPINIH 490

Search completed: January 3, 2003, 14:28:20
 Job time : 29 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 14:25:12 ; Search time 11 Seconds
(without alignments)
45.247 Million cell updates/sec

Title: US-09-899-376-1
Perfect score: 12
Sequence: 1 TSPLNHNGOKL 12

Scoring table: Oligo
Gapop 60.0 , Gapext 60.0

Searched: . 112892 seqs, 41476328 residues
Word size : 0

Total number of hits satisfying chosen parameters: 112892
Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries
Database : SwissProt;40;*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	7	58.3	321 1 Z177_HUMAN	Q13360 homo sapien
2	7	58.3	471 1 FBX3_HUMAN	Q9u99 homo sapien
3	6	50.0	294 1 YOGI_BACSU	P46340 bacillus su
4	6	50.0	391 1 Y531_METJA	Q57954 methanococc
5	6	50.0	1273 1 YAR2_SCPO	Q10135 schizosacch
6	5	41.7	111 1 YJCB_YEAST	P47062 saccharomy
7	5	41.7	113 1 SENAL_APCLA	P29233 aplysis cal
8	5	41.7	124 1 SPH2_SULSO	Q9uvyy8 sulfobolus
9	5	41.7	126 1 PX1B_CANNA	Q00680 candida mal
10	5	41.7	126 1 FX18_CANTR	P22009 candida tro
11	5	41.7	137 1 CPC_CU CSA	P29602 cucumis sat
12	5	41.7	155 1 IL17_HUMAN	Q16552 homo sapien
13	5	41.7	188 1 UBCX_PICAN	Q60015 picchia angu
14	5	41.7	193 1 FGCL_PIG	Q92505 sus scrofa
15	5	41.7	194 1 MINC_HELPJ	Q9m51 helicobacte
16	5	41.7	194 1 PGCL_MOUSE	Q00264 homo sapien
17	5	41.7	194 1 PGCL_MOUSE	Q5022 mus musculu
18	5	41.7	194 1 PGCL_RAT	P70580 rattus norv
19	5	41.7	195 1 MINC_HELP	Q25693 helicobacte
20	5	41.7	208 1 YEAF_SCHPO	Q1080 schizosacch
21	5	41.7	228 1 TFP2_PLAFA	P32086 plasmidom
22	5	41.7	241 1 CCP3_HUMAN	Q9ukab homo sapien
23	5	41.7	251 1 KDKA_VIBCH	Q9kv99 vibrio chol
24	5	41.7	265 1 YGT2_YEAST	P52262 saccharomy
25	5	41.7	266 1 KLRB_MOUSE	Q61682 mus musculu
26	5	41.7	273 1 ILVE_THEME	P74921 thermotoga
27	5	41.7	277 1 MYOP_BOVIN	P20456 bos taurus
28	5	41.7	277 1 MYOP_MOUSE	P2218 homo sapien
29	5	41.7	277 1 MYOP_RAT	Q55023 mus musculu
30	5	41.7	277 1 VPA0_MABVM	P91697 rattus norv
31	5	41.7	283 1 MYOP_XENLA	P32260 marburg vir
32	5	41.7	285 1 MYOP_XENLA	P29219 xenopus lae
33				Q9jlu8 cavia porce

ALIGNMENTS

RESULT 1	SEQUENCE FROM N.A.	RP	SEQUENCE FROM N.A.	RP
Z177_HUMAN	Z177_HUMAN	ID	Z177_HUMAN	ID
STANDARD;		CC	STANDARD;	CC
013360;		CC	013360;	CC
01-NOV-1997 (Rel. 35, Created)		CC	01-NOV-1997 (Rel. 35, Last sequence update)	CC
15-JUN-2002 (Rel. 41, Last annotation update)		DT	15-JUN-2002 (Rel. 41, Last annotation update)	DT
Zinc finger protein 177.		DE	Zinc finger protein 177.	DE
ZNF177.		GN	ZNF177.	GN
Homo sapiens (Human).		OS	Homo sapiens (Human).	OS
Eukaryota; Metazoa; Chordata; Craniata; vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.		OC	Eukaryota; Metazoa; Chordata; Craniata; vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	OC
NCBI_TaxID=9606;		OX	NCBI_TaxID=9606;	OX
[1]				
Alu and endogenous retroviral segments";		RT	Alu and endogenous retroviral segments";	RT
Genomics 33:433-472 (1995).		RL	Genomics 33:433-472 (1995).	RL
- - - FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.		RR	- - - FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.	RR
- - - SUBCELLULAR LOCATION: Nuclear (Probable).		RR	- - - SUBCELLULAR LOCATION: Nuclear (Probable).	RR
"Transcripts from a novel human KRAB zinc finger gene contain spliced		RR	"Transcripts from a novel human KRAB zinc finger gene contain spliced	RR
PINGER PROTEINS.		RR	PINGER PROTEINS.	RR
- - - SIMILARITY: CONTAINS 1 KRAB DOMAIN.		RR	- - - SIMILARITY: CONTAINS 1 KRAB DOMAIN.	RR
This SWISS-PROT entry is copyright. It is produced through a collaboration		CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	CC
between the Swiss Institute of Bioinformatics and the EMBL outstation -		CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	CC
the European Bioinformatics Institute. There are no restrictions on its		CC	use by non-profit institutions as long as its content is in no way	CC
modified and this statement is not removed. Usage by and for commercial		CC	modified and this statement is not removed. Usage by and for commercial	CC
entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to licensee@isb-sib.ch).		CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to licensee@isb-sib.ch).	CC
EMBL: U37263; AAB09749.1; -		DR	EMBL: U37263; AAB09749.1; -	DR
HSSP: P08047; ISP2.		DR	HSSP: P08047; ISP2.	DR
GeneID: 12966; ZNF177.		DR	GeneID: 12966; ZNF177.	DR
MIM: 601276; -		DR	MIM: 601276; -	DR
Inter-Pro: IPR001909; KRAB.		DR	Inter-Pro: IPR001909; KRAB.	DR
Inter-Pro: IPR00822; Znf_C2H2.		DR	Inter-Pro: IPR00822; Znf_C2H2.	DR
Pfam: PF00096; zf-f2H2; 7.		DR	Pfam: PF00096; zf-f2H2; 7.	DR
PFAM: PF01352; KRAB_1.		DR	PFAM: PF01352; KRAB_1.	DR
PRINTS: PR00048; ZINC FINGER.		DR	PRINTS: PR00048; ZINC FINGER.	DR
ProDom: PR000003; zf_f2H2; 7.		DR	ProDom: PR000003; zf_f2H2; 7.	DR
SMART: SM03349; KRAB.		DR	SMART: SM03349; KRAB.	DR
SMART: SM0355; Znf_C2H2; 7.		DR	SMART: SM0355; Znf_C2H2; 7.	DR
PROSITE: PS50805; KRAB; 1.		DR	PROSITE: PS50805; KRAB; 1.	DR
PROSITE: PS00028; ZINC FINGER_C2H2_1; 7.		DR	PROSITE: PS00028; ZINC FINGER_C2H2_1; 7.	DR
PROSITE: PS50157; ZINC FINGER_C2H2_2; 7.		DR	PROSITE: PS50157; ZINC FINGER_C2H2_2; 7.	DR
KW				
Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;				
Nuclear protein; Repeat.				
KRAB.				
ZINC FINGERS.				
C2H2-TYPE.				
C2H2-TYPE.				

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: Z68897; CAA93114.1; -
DR InterPro: IPR00937; Amino_oxidase.
DR InterPro: IPR00910; HMG_12_box.
DR Pfam: PF01593; Amino_oxidase; 1.
DR SMART: SM00398; HMG; 1.
DR Hypothetical protein; transmembrane.
FT TRANSMEM 1028 1048 POTENTIAL.
SEQUENCE 1273 AA; 142489 MW; 3FFEF643DF3C0B6 CRC64;
SQ

Query Match 50.0%; Score 6; DB 1; Length 1273;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2 SPNLN 7
Db 790 SPNLN 795

RESULT 6
YJC8 YEAST STANDARD; PRT; 111 AA.
ID P47052;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 12.9 kDa protein in MAD2-RNR2 intergenic region.
GN YJL020W OR J1267.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Saccharomyces; Saccharomycetaceae; Saccharomyces.
NCBI_TAXID=4932;
RN [1]
RR SEQUENCE FROM N.A.
RA Pohl T.M., Aljinovic G.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBDJ databases.
CC -!
CC SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Z49503; CAA89319.1; -
DR SGD; S0003565; YJL020W.
DR Hypothetical protein; transmembrane.
FT TRANSMEM 29 49 POTENTIAL.
FT TRANSMEM 52 72 POTENTIAL.
SEQUENCE 111 AA; 12897 MW; 610024D82C821589 CRC64;
SQ

Query Match 41.7%; Score 5; DB 1; Length 111;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 TSPLN 5
Db 24 TSPLN 28

RESULT 7
SENA_APCLA STANDARD; PRT; 113 AA.
ID SENA_APCLA
FT SEQUENCE 111 AA; 12897 MW; 610024D82C821589 CRC64;
SQ

Query Match 41.7%; Score 5; DB 1; Length 111;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 TSPLN 5
Db 24 TSPLN 28

RESULT 8
SPH2_SULSO STANDARD; PRT; 124 AA.
ID SPH2_SULSO
AC Q9WWT8
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE S-adenosylmethionine decarboxylase proenzyme 2 (EC 4.1.1.50) (AdoMetDC
DE 2) (SAMD2) [contains: S-adenosylmethionine decarboxylase 2 beta
DE chain; S-adenosylmethionine decarboxylase 2 alpha chain].
GN SPHE2 OR SSO085.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TAXID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
MEDLINE=2332296; PubMed=1427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awayer M.J., Chan-Waiher C.-K., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,

RA Heikamp-de Jong I., Jeffries A.C., Kozena C.J., Medina N., Peng X., RA Thi-Ngoc H.P., Reeder P., Scheink M.E., Merikant C., Tolstrup N., RX CHARLEROIS R.L., Doolittle W.F., Duguet M., Gaasterland T., RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J., Van der T., RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.;" RL proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).

[2] CHARACTERIZATION.

RT STRAIN=ATCC 49355 / DSM 5833 / MT-4;

RX MEDLINE=9130150; PubMed=1649051;

RA Cappiello G., Porcelli M., De Rosa M., Gambacorta A., Bertoldo C., RA "S'-adenosylmethionine decarboxylase from the thermophilic archaeabacterium Sulfolobus solfataricus. Purification, molecular properties and studies on the covalently bound pyruvate.;"

RL RT properties on the covalently bound pyruvate.;"

CC RT FUNCTION: DECARBOXYLATION OF S'ADENOSYLMETHIONINE PROVIDES THE AMINOPROPYL MOIETY REQUIRED FOR SPERMIDINE BIOSYNTHESIS FROM PUTRESCINE (BY SIMILARITY).

CC RT -I- CATALYTIC ACTIVITY: S'-adenosyl-L-methionine = (5-deoxy-5-adenylyl)(3-aminopropyl) methylsulfonium salt + CO(2).

CC RT -I- COFACTOR: Pyruvyl group.

CC RT -I- MISCELLANEOUS: THERMOPHILIC AND THERMOSTABLE.

CC RT -I- SIMILARITY: BELONGS TO THE PROKARYOTIC ADOMETDC FAMILY. SUBFAMILY 1.

CC RT -----

CC RT This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC RT -----

CC DR EMBL: M61102; AA343359.1; -.

CC DR IPR00590; JT0590.

CC DR InterPro; IPR003033; SCP2.

CC DR Pfam; PF02036; SCP2; 1.

KW Lipid-binding; Peroxisome.

FT INIT_MET 0 BY SIMILARITY.

FT DOMAIN 32 40 HYDROPHOBIC.

FT DOMAIN 42 51 HYDROPHILIC.

SQ SEQUENCE 126 AA: 13687 MW: 385d82d357ebe8fe CRC64;

Query Match 41.7%; Score 5; DB 1; Length 126; Best Local Similarity 100.0%; Pred. No. 36; Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0; QY 8 NGOKL 12

Db 91 NGOKL 95

RESULT 10

PX18_CANTR STANDARD; PRT; 126 AA.

ID PX18_CANTR STANDARD; PRT; 126 AA.

AC P22009;

DT 01-AUG-1991 (Rel. 19, Created)

DT 01-MAY-1992 (Rel. 22, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Oleate-induced peroxisomal protein Pox18 (Lipid-transfer protein (PXP-18)).

DE Pox18.

OS Candida tropicalis (Yeast).

OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; OC Saccharomycetales; mitosporic Saccharomycetales; Candida.

OX NCBI_TAXID=5482;

RN [1]

SEQUENCE FROM N.A., AND SEQUENCE OF 1-16.

RP STRAIN=TCC 20336 / PK233;

RX MEDLINE=89252907; PubMed=2470648;

RA Szabo L.J., Small G.M., Lazarow P.B.;

RT "The nucleotide sequence of Pox18, a gene encoding a small oleate-inducible peroxisomal protein from Candida tropicalis.;"

RT Gene 75:119-126(1989).

RN [2]

SEQUENCE FROM N.A.

RP STRAIN=ATCC 20336 / PK233;

RC MEDLINE=90305990; PubMed=2364939;

RA Tan H., Okazaki K., Kubota I., Kamiryo T., Utivama H.;

RT "A novel peroxisomal nonspecific lipid-transfer protein from Candida tropicalis. Gene structure, purification and possible role in beta-

RN [1]

CC INTRACELLULAR ADHESION MOLECULE-1 (ICAM-1) IN FIBROBLASTS.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: RESTRICTED TO ACTIVATED MEMORY T-CELLS.
 CC -!- PTM: FOUND BOTH IN GLYCOSYLATED AND NONGLYCOSYLATED FORMS.
 CC -!- SIMILARITY: BELONGS TO THE IL-17 FAMILY.
 CC -!- DATABASE: NAME=REO System' cytokine mini-reviews: IL17;
 CC WWW="http://www.rndsystems.com/asp/g_sitebuilder.asp?bodyId=211".
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: Z58820; CRA91233; 1; -.
 DR EMBL; U32659; AAC50341; 1; -.
 DR Genew; HGNC: 5981; IL17.
 DR
 DR
 KW Cytokine; Glycoprotein; T-cell; Antigen; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 155
 FT BY SIMILARITY.
 FT DISULFID 94 144
 FT DISULFID 99 146
 FT CARBOHYD 68 68
 SQ SEQUENCE 155 AA; 17504 MW; 2BCAB9CB2F4886D1 CRC64;
 Query Match 41.7%; Score 5; DB 1; Length 155;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 4 UNIH 8
 Db 49 UNIH 53
 RESULT 13
 ID UBCX_PICAN STANDARD; PRT; 188 AA.
 AC 060015;
 DT 15-DEC-1998 (Rel. 37; Created)
 DT 15-DEC-1998 (Rel. 37; Last sequence update)
 DT 15-JUN-2002 (Rel. 41; Last annotation update)
 DE (Ubiquitin-conjugating enzyme E2-21 kDa (EC 6.3.2.19)
 DE (Ubiquitin protein ligase) (Ubiquitin carrier protein) (Peroxin-4).
 GN PEX4.
 OS Pichia angusta (Yeast) (Hansenula polymorpha).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Pichia.
 RN [1] NCBI Taxid=4905;
 RN SEQUENCE FROM N.A.
 RC STRAIN=CBS 4732;
 RC MEDLINE=98315053; PubMed=9649431;
 RA van der Klei I.J., Hilbrands R.E., Kiel J.A.K.W., Rasmussen S.W.,
 RA Cregg J.M., Veenhuus M.;
 RT "The ubiquitin-conjugating enzyme Pex4p of Hansenula polymorpha is
 RT required for efficient functioning of the PTS1 import machinery.",
 RL EMBJ J. 17:33608-33618(1998).
 CC -!- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER
 CC -!- PROTEINS. ESSENTIAL FOR PEROXISOME BIOGENESIS.
 CC -!- CATALYTIC ACTIVITY: ATP + ubiquitin + protein lysine = AMP +
 CC diphosphate + protein N-ubiquityllysine.
 CC -!- PATHWAY: Ubiquitin conjugation; second step.
 CC -!- MISCCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-
 CC THIOLESTER FORMATION.
 CC -!- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X99714; CAA68050; 1; -.
 DR InterPro; IPR01199; Cyt_B5.
 DR Pfam; PF00173; heme_1; 1.
 KW Receptor; Steroid binding; Transmembrane; Microsome.
 FT INIT_MET 0

F^{RT} TRANSMEM 24 42 POTENTIAL.
 SQ SEQUENCE 193 AA; 21478 MW; F7524857C95CDB23 CRC64;
 Query Match 41.7%; Score 5; DB 1; Length 193;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TAPLN 5
 1111
 26 TSPLN 30

RESULT 15

MNC_HELPJ STANDARD; PRT; 194 AA.

ID MNC_HELPJ STANDARD; PRT; 194 AA.

AC QP2M51;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

DE Probable septum site-determining protein minC.

GN MNC OR JHP0372.

OS Helicobacter pylori J99 (Campylobacter pylori J99);

OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;

OC Helicobacter.

OX NCBI_TAXID=85963;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99120557; PubMed=9923682;

RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,

RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,

RA Tummino P.J., Caruso A., Urias-Nickelsen M., Mills D.M., Ives C.,

RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,

RA Trust T.J.;

RT "Genomic sequence comparison of two unrelated isolates of the human

RT gastric pathogen Helicobacter pylori.";

RL Nature 397:176-180(1999).

CC FUNCTION: CELL DIVISION INHIBITOR THAT BLOCKS THE FORMATION OF

CC POLAR 2 RING SEPTA. RAPIDLY OSCILLATES BETWEEN THE POLES OF THE

CC CELL TO DESTABILIZE FTSZ FILAMENTS THAT HAVE FORMED BEFORE THEY

CC MATURE INTO POLAR 2 RINGS. PREVENT FTSZ POLYMERIZATION (BY

CC SIMILARITY).

CC -!- SUBUNIT: INTERACTS WITH MIND AND FTSZ (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE MNC FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

CC

DR EMBL: AE001472; ADD:0953 1; -

DR KW Cell division; Septation; Complete proteome.

SQ SEQUENCE 194 AA; 22302 MW; 33093F48637D0FAB CRC64;

Query Match 41.7%; Score 5; DB 1; Length 194;

Best Local Similarity 100.0%; Pred. No. 54;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 NTHNG 9
 1111
 117 NTHNG 121

Search completed: January 3, 2003, 14:27:02

Job time : 11 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 14:26:33 ; Search time 16 Seconds

Sequence: (without alignments) 72.101 Million cell updates/sec

Title: US-09-899-376-1
Perfect score: 12
Sequence: 1 TSPLNIHNGOKL 12

Scoring table: OLIGO Gapext 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB ID	Description
1	6	50.0	169	2	G83075	type 4 fimbrial bi
2	6	50.0	294	2	C69956	phosphate ABC tran
3	6	50.0	391	2	F64366	flavoprotein - Met
4	6	50.0	481	2	T32260	hypothetical prote
5	6	50.0	706	2	S42174	NPtase - African s
6	6	50.0	706	2	JQ2210	probable nucleic a
7	6	50.0	1050	2	T18300	serine/threonine-s
8	6	50.0	1273	2	T38292	hypothetical prote
9	6	50.0	1690	2	T31670	DNA-directed RNA p
10	5	41.7	61	2	AD2424	hypothetical prote
11	5	41.7	63	2	D81307	hypothetical prote
12	5	41.7	69	2	D97809	hypothetical prote
13	5	41.7	76	2	T17971	hypothetical prote
14	5	41.7	80	2	S77059	hypothetical prote
15	5	41.7	100	2	B71102	hypothetical prote
16	5	41.7	111	2	S56802	probable membrane
17	5	41.7	113	2	S23653	sensorin A - Calif
18	5	41.7	124	2	H87616	conserved hypothet
19	5	41.7	127	2	T44119	cytochrome c famil
20	5	41.7	127	2	JS0155	hypothetical prote
21	5	41.7	127	2	JT0590	lipid transfer pro
22	5	41.7	127	2	JT0590	POX18Cm protein -
23	5	41.7	128	2	C84708	probable MYB famil
24	5	41.7	128	2	C90205	hypothetical prote
25	5	41.7	135	2	H71051	hypothetical prote
26	5	41.7	137	1	SSKV	cupredoxin (valida
27	5	41.7	139	2	A13059	hypothetical prote
28	5	41.7	139	2	G98226	hypothetical prote
29	5	41.7	145	2	B96705	unknown protein, 8

ALIGNMENTS

RESULT 1
G83075 type 4 fimbrial biogenesis protein FimT PA4549 [imported] - *Pseudomonas aeruginosa* (s

C;Species: *Pseudomonas aeruginosa*
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: G83075

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; admian, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L; i; Iorly, S.; Olson, M.V.

Nature 405, 959-964, 2000

A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: G83075

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-169 <STO>

A;Cross-references: GB:AE0048669; GB:AE004091; NID:9950793; PIDN:AAG07937.1; GSPDB:GN

A;Experimental source: strain PA01

C;Genetics: fimb; PA4549

A;Gene: fimb; PA4549

Query	Match	Length	DB ID	Description
6	4	LNHNG	9	type 4 fimbrial bi
6	1	11111	123	phosphate ABC tran
6	1	11111	128	phosphate ABC tran

RESULT 2
C69956

Query Match

50.0%; Score 6; DB 2;

Length 169;

Best Local Similarity 100.0%; Pred No. 8.5; Mismatches 0; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

4

LNHNG

9

Db

123

LNIHNG

128

C;Species: *Bacillus subtilis*

C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000
C;Accession: C69956

R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Ellermeier, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal

iech, J.; Harwood, C.R.; Renaut, A.; Hibert, H.; Holsappel, S.; Hosono, S.; Hullo, M

Koettler, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

Rieger, M.; Rivolta, C.; Rocha, E.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl

A;Authors: Schleicher, S.; Schreiter, R.; Scorfone, P.; Sekiguchi, J.; Sekowska, A.; Se

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Togoni, A.; Tostao, V.; Uchida

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, S.; Yoshida

A;Authors: Yoshikara, H.F.; Zumstein, E.; Yoshioka, H.; Danchin, A.

A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*

A;Reference number: A669580; MUID:98044033; PMID:9384377

				F;340-420/Domain: ets DNA-binding domain homology <ET5>
A;Accession:	C69956			
A;Status:	preliminary; nucleic acid sequence not shown; translation not shown			
A;Molecule type:	DNA			
A;Residues:	1-294 <KUN>			
A;Cross-references:	GB:299116; GB:AL009126; NID:92634723; PIDN:CAB1427.1; PID:92634930			
A;Experimental source:	strain 168			
A;Gene:	yqgl			
C;Superfamily:	pho protein			
Query Match	50.0%; Score 6; DB 2; Length 294;			
Best Local Similarity	100.0%; Pred. No. 14;			
Matches	6; Conservative 0; Mismatches 0;			
Indels	0;			
Gaps	0;			
RESULT 3				
flavoprotein - Methanococcus jannaschii				
C;Species:	Methanococcus jannaschii			
C;Date:	13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 18-Sep-1998			
C;Accession:	F64366			
R;Blatt, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Carson, J.D.; Sadow, P.W.; Hama, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.				
Science 273, 158-167, 1996				
A;Authors:	Kane, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Freije, J.M.P.; Lain, S.; Vinuela, E.; Lopez-Otin, C.			
A;Title:	Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii			
A;Reference number:	A64300; MUID:96337999; PMID:8688087			
A;Status:	preliminary; nucleic acid sequence not shown; translation not shown			
A;Molecule type:	DNA			
A;Residues:	1-391 <BUL>			
A;Cross-references:	GB:U67502; GB:L77117; NID:91591223; PID:91591237; TIGR:M0534; PID:91591238			
C;Genetics:				
A;Map position:	FOR469119-470294			
A;Start codon:	GTC			
C;Superfamily:	Methanobacterium flavoprotein A			
C;Keywords:	flavoprotein			
Query Match	50.0%; Score 6; DB 2; Length 391;			
Best Local Similarity	100.0%; Pred. No. 18;			
Matches	6; Conservative 0; Mismatches 0;			
Indels	0;			
Gaps	0;			
Qy 1 TSPN1 6				
Db 283 TSPN1 288				
RESULT 4				
hypothetical protein C24A1.2 - Caenorhabditis elegans				
C;Species:	Caenorhabditis elegans			
C;Date:	29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 02-Sep-2000			
C;Accession:	T32260			
R;Connell, M.				
submitted to the EMBL Data Library, September 1997				
A;Description:	The sequence of C. elegans cosmid C24A1.			
A;Reference number:	221141			
A;Accession:	T32260			
A;Status:	preliminary; translated from GB/EMBL/DBJ			
A;Molecule type:	DNA			
A;Residues:	1-481 <CON>			
A;Cross-references:	EMBL:AF024491; PIDN:AAB70311.1; GSPDB:GN00021; CESP:C24A1.2			
A;Experimental source:	strain Bristol N2; clone C24A1			
A;Gene:	CESP:C24A1.2			
A;Map position:	3			
A;Introns:	55/3; 96/1; 129/2; 190/2; 209/3; 233/1; 277/1; 428/3; 445/3			
C;Superfamily:	ets DNA-binding domain homology			
Query Match	50.0%; Score 6; DB 2; Length 706;			
Best Local Similarity	100.0%; Pred. No. 31;			
Matches	6; Conservative 0; Mismatches 0;			
Indels	0;			
Gaps	0;			
Qy 2 SPLNH 7				
Db 485 SPLNH 490				
RESULT 7				
T18300				

serine/threonine-specific protein kinase homolog - *Emericella nidulans*
 C;Species: *Emericella nidulans*, *Aspergillus nidulans*
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C;Accession: T18300
 R;Ramos, P.; Ellis, L.
 R;Submitted to the EMBL Data Library, April 1995
 A;Reference number: Z18688
 A;Accession: T18300
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-1050 <RAN>
 A;Cross-references: EMBL:U25693; NID:q818862; PID:q818863; PIDN:AA67069.1
 A;Introns: 626/1; 808/3
 A;Note: anKA

RESULT 8
 T38292 hypothetical protein SPAC23E2.02 - fission yeast (*Schizosaccharomyces pombe*)
 C;Species: *Schizosaccharomyces pombe*
 C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C;Accession: T18292
 R;Skelton, J.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
 R;Submitted to the EMBL Data Library, January 1996
 A;Reference number: Z21784
 A;Accession: T18292
 A;Molecule type: DNA
 A;Residues: 1-1273 <SKE>
 A;Cross-references: EMBL:Z68887; PIDN:CAA93114.1; GSPDB:GN00066; SPDB:SPAC23E2.02
 A;Experimental source: strain 972h-; cosmid c23E2
 C;Genetics:
 A;Gene: SPDB:SPAC23E2.02
 A;Map position: 1
 A;Introns: 8/2; 862/3

Query Match 50.0%; Score 6; DB 2; Length 1050;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 6; Conservative 0; Mismatches 0;
 Qy 3 PLNIN 8
 Db 661 PLNIN 666

RESULT 9
 T31670 DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain - *Mastigamoeba invertens* (frag
 C;Species: *Mastigamoeba invertens*
 C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 03-Nov-2000
 C;Accession: T31670
 R;Stiller, J.W.; Duffield, E.C.S.; Hall, B.D.
 R;Proc. Natl. Acad. Sci. U.S.A. 95, 11769-11774, 1998
 A;Title: Amitochondriate amoebae and the evolution of DNA-dependent RNA polymerase II.
 A;Reference number: Z21052; MVID:98426228; PMID:9751740
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-1690 <ST1>
 A;Cross-references: EMBL:AF083338; NID:q3511286; PID:q3511287; PIDN: AAC62246.1
 A;Experimental source: ATCC50338
 C;Genetics:
 A;Gene: RPBI

A;Note: Intron positions not resolved (incomplete sequence)

RESULT 10
 AD2424 hypothetical protein asl4948 [imported] - *Nostoc* sp. (strain PCC 7120)
 C;Species: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
 C;Accession: AD2424
 C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
 R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kiritz, T.; Sasamoto, S.; Watanabe, A.; Iriu
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
 DNA Res. 8, 205-213, 2001
 A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
 A;Reference number: AD1807; MVID:2195285; PMID:11759840
 A;Accession: AD2424
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-61 <KUR>
 A;Cross-references: GB:BA000019; PIDN:BA876647.1; PID:917134086; GSPDB:GN00179
 A;Experimental source: strain PCC 7120
 C;Genetics:
 A;Gene: asl4948

RESULT 11
 B81307 hypothetical protein CJ1047c [imported] - *Campylobacter jejuni* (strain NCTC 11168)
 C;Species: *Campylobacter jejuni*
 C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
 C;Accession: D81307
 R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanTiel, A.; Whitehead, S.; Bar
 Nature 403, 665-668, 2000
 A;Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals
 A;Reference number: A81250; MVID:20150912; PMID:10688204
 A;Accession: D81307
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-63 <PAR>
 A;Cross-references: GB:AU139077; GB:AU11168; NID:96968444; PIDN:CAB73303.1; PID:9696
 A;Experimental source: serotype O2, strain NCTC 11168
 C;Genetics:
 A;Gene: CJ1047c

RESULT 12
 D97809

hypothetical protein RC0876 [imported] - *Rickettsia conorii* (strain Malish 7)
 C;Species: *Rickettsia conorii*
 C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
 C;Accession: D97809
 R;Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Roc
 Science, 293, 2093-2098, 2001
 A;Title: Mechanisms of evolution in *Rickettsia conorii* and *Rickettsia prowazekii*.
 A;Reference number: A97700; MUID:21442074; PMID:11557893
 A;Accession: D97809
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-69 <KUR>
 A;Cross-references: GB:AE006914; PIDN:AAL03414.1; PID:915619981; GSPDB:GN00173
 A;Gene: RC0876
 Query Match 41.7%; Score 5; DB 2; Length 69;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 7 HNGQK 11
 |||||
 Db 23 HNGQK 27

RESULT 13

T17971 hypothetical protein a469L - *Chlorella virus* PBCV-1
 C;Species: *Chlorella virus* PBCV-1
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C;Accession: T17971
 R;Graves, M.V.; Van Etten, J.L.; submitted to the EMBL Data Library, May 1999
 A;Reference number: 218806
 A;Accession: T17971
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-76 <GRA>
 A;Cross-references: EMBL:U42580; NID:94028896; PIDN: AAC96836.1
 A;Experimental source: specific host *Chlorella* strain NC64A
 C;Genetics:
 A;Note: a469L

Query Match 41.7%; Score 5; DB 2; Length 76;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 TSPLN 5
 |||||
 Db 56 TSPLN 60

RESULT 14

S77059 hypothetical protein ssi1263 - *Synechocystis* sp. (strain PCC 6803)
 C;Species: *Synechocystis* sp.
 A;Variety: PCC 6803
 C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
 C;Accession: S77059
 R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Oka, K.; Okumura, S.; Shimojo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, O.; DNA Res. 3, 109-136, 1995
 A;Title: sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp.
 A;Reference number: S74322; MUID:97061201; PMID:8905231
 A;Accession: S77059
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-80 <KAN>
 A;Cross-references: EMBL:D64005; GB:AB001339; NID:91001779; PIDN:BAA10761.1; PID:d101141
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 41.7%; Score 5; DB 2; Length 80;

Best Local Similarity 100.0%; Pred. No. 57;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSPLN 5
 |||||
 Db 60 TSPLN 64

RESULT 15

B71102

hypothetical protein PH1077 - *Pyrococcus horikoshi*

C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000

C;Accession: B71102
 R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
 M.; Ohfuki, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Ogu
 DNA Res. 5, 55-76, 1998

A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic

A;Reference number: A71000; MUID:98344137; PMID:9679194

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA
 A;Residues: 1-100 <RAW>

A;Cross-references: GB:AP000004; NID:93236131; PIDN:BA30176.1; PID:93257493

A;Experimental source: strain On3
 A;Note: This accession replaces an interim accession for a sequence replaced by GenBank

C;Genetics: PH1077

C;Superfamily: *Pyrococcus horikoshi* hypothetical protein PH1077

Query Match 41.7%; Score 5; DB 2; Length 100;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSPLN 5
 |||||
 Db 83 TSPLN 87

Search completed: January 3, 2003, 14:28:42
 Job time : 17 secs

GenCore Version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 14:24:57 ; Search time 11 Seconds
(without alignments)
20.674 Million cell updates/sec

Title: US-09-899-376-1

Perfect score: 64

Sequence: 1 TSPLNITHNGQKL 12

Scoring table: BLOSUM2

Gapop 10.0 , Gapext 0.5

Searched: 117078 seqs, 18951520 residues

Total number of hits satisfying chosen parameters: 117078

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB_pep:*

2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB_pep:*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB_pep:*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB_pep:*

5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB_pep:*

6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB_pep:*

7: /cgn2_6/ptodata/1/pubpaa/PCVUS_PUBCOMB_pep:*

8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB_pep:*

9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB_pep:*

10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB_pep:*

11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB_pep:*

12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB_pep:*

13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB_pep:*

14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB_pep:*

RESULT 1
US-09-899-376-1
; Sequence 1, Application US/09899376
; Patent No. US20020102265A1
; GENERAL INFORMATION:
; APPLICANT: HONG, FRANK D.
; ATTORNEY: CLAYMAN, GARY
; TITLE OF INVENTION: ISOLATION OF A CELL-SPECIFIC INTERNALIZING PEPTIDE THAT INFILTRATES TUMOR TISSUE FOR TARGETED DRUG DELIVERY
; FILE REFERENCE: UTSC:645US
; CURRENT APPLICATION NUMBER: US/09/899,376
; CURRENT FILING DATE: 2001-07-02
; NUMBER OF SEQ ID NOS: 5
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-09-899-376-1
Query Match 100.0%; Score 64; DB 10; Length 12;
Best Local Similarity 100.0%; Pred. No. 3.3e-06;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2
US-09-899-376-3
; Sequence 3, Application US/09899376
; Patent No. US20020102265A1
; GENERAL INFORMATION:
; APPLICANT: HONG, FRANK D.
; ATTORNEY: CLAYMAN, GARY
; TITLE OF INVENTION: ISOLATION OF A CELL-SPECIFIC INTERNALIZING PEPTIDE THAT INFILTRATES TUMOR TISSUE FOR TARGETED DRUG DELIVERY
; FILE REFERENCE: UTSC:645US

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64	100.0	12	10	US-09-899-376-1
2	64	100.0	19	10	US-09-899-376-3
3	64	100.0	19	10	US-09-899-376-4
4	42	1325	10	US-09-741-669-304	
5	40	62.5	61	10	US-09-864-761-40714
6	38	59.4	247	12	US-10-042-417-36
7	37	57.8	240	10	US-09-815-242-5066
8	36	56.2	194	10	US-09-815-242-11500
9	36	56.2	196	10	US-09-815-242-11660
10	36	56.2	252	10	US-09-815-242-10296
11	36	56.2	259	10	US-09-864-761-13738
12	35	54.7	86	10	US-09-864-761-39192
13	35	54.7	160	10	US-09-864-761-47255
14	35	54.7	1088	10	US-09-920-804-2
15	35	54.7	2000	12	US-10-010-901-29
16	35	54.7	2150	9	US-10-130-322-17
17	34	53.1	287	10	US-09-729-714
18	34	53.1	287	10	US-09-800-729-153
19	34	53.1	287	10	US-09-925-297-786
20	34	53.1	478	10	US-09-417-791-5
21	34	53.1	558	10	US-09-864-761-47724
22	34	53.1	613	10	US-09-800-729-82
23	34	53.1	613	10	US-09-800-729-98
24	33	51.6	33	10	US-09-864-761-34602
25	33	51.6	209	10	US-09-761-36708
26	33	51.6	209	10	US-09-764-853-546
27	33	51.6	278	10	US-09-903-456-18
28	33	51.6	337	9	US-10-063-547-74
29	33	51.6	337	12	US-10-005-867-74
30	33	51.6	337	12	US-10-052-586-268
31	33	51.6	342	10	US-09-861-761-12637
32	33	51.6	345	10	US-09-815-242-11499
33	33	51.6	505	10	US-09-815-242-5857
34	33	51.6	505	10	US-09-811-242-12868
35	33	51.6	870	10	US-09-815-242-5493
36	33	51.6	870	10	US-09-815-242-12637
37	33	51.6	2139	10	US-09-727-384-6
38	33	51.6	2697	10	US-09-961-527A-5
39	33	51.6	3571	9	US-10-150-821-2
40	33	51.6	3571	10	US-09-911-842-2
41	32.5	50.8	677	10	US-09-815-242-11921
42	32	50.0	10	US-09-984-056-11	
43	32	50.0	10	US-09-984-057-11	
44	32	50.0	30	10	US-09-989-789-9
45	32	50.0	30	10	US-09-989-789-9
46	50.0	50.0	60	10	US-09-764-877-1853

ALIGNMENTS

Sequence 5, App1
Sequence 18, App1
Sequence 74, App1
Sequence 268, App1
Sequence 36708, App1
Sequence 11499, App1
Sequence 587, App1
Sequence 12868, App1
Sequence 593, App1
Sequence 12637, App1
Sequence 6, App1
Sequence 5, App1
Sequence 2, App1
Sequence 11921, App1
Sequence 11, App1
Sequence 11, App1
Sequence 9, App1
Sequence 1853, App1

CURRENT APPLICATION NUMBER: US/09/899,376 ;
 CURRENT FILING DATE: 2001-07-02 ;
 NUMBER OF SEQ ID NOS: 5 ;
 SOFTWARE: PatentIn ver. 2.1
 SEQ ID NO 3 LENGTH: 19
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence:
 OTHER INFORMATION: Peptide

Query Match 100.0%; Score 64; DB 10; Length 19;
 Best Local Similarity 100.0%; Pred. No. 5;e-06; Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 GENERAL INFORMATION:
 APPLICANT: HONG, FRANK D.
 APPLICANT: CLAYMAN, GARY
 TITLE OF INVENTION: ISOLATION OF A CELL-SPECIFIC INTERNALIZING PEPTIDE THAT INFILTRATES TUMOR TISSUE FOR TARGETED DRUG DELIVERY
 TITLE OF INVENTION: FILE REFERENCE: US/09/6450S
 CURRENT APPLICATION NUMBER: US/09/899,376
 CURRENT FILING DATE: 2001-07-02
 NUMBER OF SEQ ID NOS: 5
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 4 LENGTH: 19
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: Synthetic

Query Match 100.0%; Score 64; DB 10; Length 19;
 Best Local Similarity 100.0%; Pred. No. 5;e-06; Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 GENERAL INFORMATION:
 APPLICANT: FORSYTH, R. ALYN
 APPLICANT: OHLSEN, KARL L.
 APPLICANT: ZYRSKIN, JUDITH W.
 TITLE OF INVENTION: Genes identified as required for proliferation of E. coli
 FILE REFERENCE: ELTRIA 009A
 CURRENT APPLICATION NUMBER: US/09/741,569
 CURRENT FILING DATE: 2000-12-19
 PRIOR APPLICATION NUMBER: US 60/173005
 PRIOR FILING DATE: 1999-12-23
 NUMBER OF SEQ ID NOS: 481
 SOFTWARE: FastSEQ for Windows Version 4.0
 SEQ ID NO 34 LENGTH: 1325

RESULT 3
 US-09-899-376-4
 Sequence 4, Application US/09899376
 Patent No. US20020102265A1
 GENERAL INFORMATION:
 APPLICANT: HONG, FRANK D.
 APPLICANT: CLAYMAN, GARY
 TITLE OF INVENTION: ISOLATION OF A CELL-SPECIFIC INTERNALIZING PEPTIDE THAT INFILTRATES TUMOR TISSUE FOR TARGETED DRUG DELIVERY
 TITLE OF INVENTION: FILE REFERENCE: US/09/6450S
 CURRENT APPLICATION NUMBER: US/09/899,376
 CURRENT FILING DATE: 2001-07-02
 NUMBER OF SEQ ID NOS: 5
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 4 LENGTH: 19
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: Synthetic

Query Match 100.0%; Score 64; DB 10; Length 19;
 Best Local Similarity 100.0%; Pred. No. 5;e-06; Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 GENERAL INFORMATION:
 APPLICANT: HONG, FRANK D.
 APPLICANT: CLAYMAN, GARY
 TITLE OF INVENTION: ISOLATION OF A CELL-SPECIFIC INTERNALIZING PEPTIDE THAT INFILTRATES TUMOR TISSUE FOR TARGETED DRUG DELIVERY
 TITLE OF INVENTION: FILE REFERENCE: US/09/6450S
 CURRENT APPLICATION NUMBER: US/09/899,376
 CURRENT FILING DATE: 2001-07-02
 NUMBER OF SEQ ID NOS: 5
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 4 LENGTH: 19
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence:
 OTHER INFORMATION: Peptide

RESULT 4
 US-09-741-669-304
 Sequence 304, Application US/09741669
 Patent No. US2002022718A1
 GENERAL INFORMATION:
 APPLICANT: FORSYTH, R. ALYN
 APPLICANT: OHLSEN, KARL L.
 APPLICANT: ZYRSKIN, JUDITH W.
 TITLE OF INVENTION: Genes identified as required for proliferation of E. coli
 FILE REFERENCE: ELTRIA 009A
 CURRENT APPLICATION NUMBER: US/09/741,569
 CURRENT FILING DATE: 2000-12-19
 PRIOR APPLICATION NUMBER: US 60/173005
 PRIOR FILING DATE: 1999-12-23
 NUMBER OF SEQ ID NOS: 481
 SOFTWARE: FastSEQ for Windows Version 4.0
 SEQ ID NO 34 LENGTH: 1325

RESULT 5
 US-09-899-376-4
 Sequence 40714, Application US/09864761
 Patent No. US2002048763A1
 GENERAL INFORMATION:
 APPLICANT: Penn, Sharron G.
 APPLICANT: Rank, David R.
 APPLICANT: Chen, Wensheng
 APPLICANT: Hanzel, David K.
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR FILE REFERENCE: Aeonica-X-1
 CURRENT APPLICATION NUMBER: US/09/864,761
 CURRENT FILING DATE: 2001-05-23
 PRIOR APPLICATION NUMBER: US 60/180,312
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: US 60/207,456
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: US 09/632,366
 PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: GB 24263-6
 PRIOR FILING DATE: 2000-10-04
 PRIOR APPLICATION NUMBER: US 60/236,359
 PRIOR FILING DATE: 2000-09-27
 PRIOR APPLICATION NUMBER: PCT/US01/00666
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
 SEQ ID NO 40714 LENGTH: 61
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO U82670_2
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.7
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.5

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
; OTHER INFORMATION: SWISSPROT HIT: P94598, EVALUATE 1.50e+00
; US-09-864-761-40714

Query Match 62.5%; Score 40; DB 10; Length 61;
; Best Local Similarity 72.7%; Pred. No. 0.6; 8; Conservative 1; Mismatches
; Matches 2; Indels 0; Gaps 0;
; Db 35 TSPKHKNGQK 45

RESULT 6
; US-10-042-417-36
; Sequence 36, Application US/10042417
; Patent No. US2002123082A1

GENERAL INFORMATION:
; APPLICANT: Pagano, M.
; TITLE OF INVENTION: METHODS TO IDENTIFY COMPOUNDS USEFUL FOR THE TREATMENT OF
; FILE REFERENCE: 5914-090-999
; CURRENT APPLICATION NUMBER: US/10/042,417
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: 60/260,179
; PRIOR FILING DATE: 2001-01-15
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-042-417-36

Query Match 59.4%; Score 38; DB 12; Length 247;
; Best Local Similarity 100.0%; Pred. No. 6.9; 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
; Matches 0; Db 146 IHNGQKL 152

RESULT 7
; US-09-815-242-5056
; Sequence 5056, Application US/09815242
; Patent No. US20081569A1

GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Travick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELTRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/259,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11500
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; US-09-815-242-11500

Query Match 55.2%; Score 36; DB 10; Length 94;
; Best Local Similarity 75.0%; Pred. No. 12; 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
; Matches 6; Db 117 NIHNGAKI 124

RESULT 9
; PRIOR APPLICATION NUMBER: 60/206,848
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11500
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; US-09-815-242-11500

Query Match 55.2%; Score 36; DB 10; Length 94;
; Best Local Similarity 75.0%; Pred. No. 12; 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
; Matches 6; Db 117 NIHNGAKI 124

US-09-815-242-11660
Sequence 11660, Application US/09815242
GENERAL INFORMATION:
PATENT NO. US20020061569A1
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815, 242
CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191, 078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206, 848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207, 727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242, 578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253, 625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257, 931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269, 308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 10296
LENGTH: 252
TYPE: PRT
ORGANISM: Escherichia coli
US-09-815-242-10296
Query Match 56.2%; Score 36; DB 10; Length 252;
Best Local Similarity 77.8%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 0;
Indels 0; Gaps 0;
Qy 2 SPLNTHHQ 10
Db 85 SPLHHLGQ 93
RESULT 11
US-09-815-242-11738
Query Match 56.2%; Score 36; DB 10; Length 252;
Best Local Similarity 77.8%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 0;
Indels 0; Gaps 0;
Qy 2 SPLNTHHQ 10
Db 85 SPLHHLGQ 93
TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815, 242
CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191, 078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206, 848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207, 727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242, 578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253, 625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257, 931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269, 308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 13738
LENGTH: 259
TYPE: PRT
ORGANISM: Salmonella typhi
US-09-815-242-13738
Query Match 56.2%; Score 36; DB 10; Length 259;
Best Local Similarity 77.8%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 0;
Indels 0; Gaps 0;
Qy 2 SPLNTHHQ 10
Db 92 SPLHHLGQ 100

RESULT 12
 US-09-864-761-39192
 Sequence 39192, Application US/09864761
 Patent No. US20020048763A1
 GENERAL INFORMATION:
 APPLICANT: Penn, Sharron G.
 APPLICANT: Rank, David R.
 APPLICANT: Hanzel, David K.
 APPLICANT: Chen, Wensheng
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
 FILE REFERENCE: Aeonica-X-1
 CURRENT APPLICATION NUMBER: US/09/864,761
 CURRENT FILING DATE: 2001-05-23
 PRIOR APPLICATION NUMBER: US 60/180,312
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: US 60/207,456
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: US 09/632,366
 PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: GB 24263.6
 PRIOR FILING DATE: 2000-10-04
 PRIOR APPLICATION NUMBER: US 60/236,359
 PRIOR FILING DATE: 2000-09-27
 PRIOR APPLICATION NUMBER: PCT/US01/00666
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
 SEQ ID NO 39192
 LENGTH: 86
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AC003684.1
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.8
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.2
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.6
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.1
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.3
 OTHER INFORMATION: SWISSPROT HIT: 070311, EVALUE 3.90e+00
 ; OTHER INFORMATION: SWISSPROT HIT: 070311, EVALUE 3.90e+00
 ; OTHER INFORMATION: SWISSPROT HIT: P52742, EVALUE 1.00e-55
 ; OTHER INFORMATION: SWISSPROT HIT: B590267.1, EVALUE 4.00e-99
 ; OTHER INFORMATION: EST-HUMAN HIT: B590267.1, EVALUE 4.00e-99

RESULT 13
 US-09-864-761-47255
 Sequence 47255, Application US/09864761
 Patent No. US20020048763A1
 GENERAL INFORMATION:
 APPLICANT: Rank, David R.
 APPLICANT: Hanzel, David K.
 APPLICANT: Chen, Wensheng
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
 FILE REFERENCE: Aeonica-X-1
 CURRENT APPLICATION NUMBER: US/09/864,761
 CURRENT FILING DATE: 2001-05-23
 PRIOR APPLICATION NUMBER: US 60/180,312
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: US 60/207,456
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: US 09/632,366
 PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: GB 24263.6
 PRIOR FILING DATE: 2000-10-04
 PRIOR APPLICATION NUMBER: US 60/236,359
 PRIOR FILING DATE: 2000-09-27
 PRIOR APPLICATION NUMBER: PCT/US01/00666
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
 SEQ ID NO 47255
 LENGTH: 160
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AC002310.1
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 13
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.52
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.53
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
 OTHER INFORMATION: SWISSPROT HIT: P52742, EVALUE 1.00e-55
 OTHER INFORMATION: EST-HUMAN HIT: B590267.1, EVALUE 4.00e-99

Query Match 54.7%; Score 35; DB 10; Length 86;
 Best Local Similarity 62.5%; Pred. No. 7.5; Mismatches 1; Indels 0; Gaps 0;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Query Match 54.7%; Score 35; DB 10; Length 160;
 Best Local Similarity 54.5%; Pred. No. 15;

```

Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Qy 1 TSPLNIHNGQK 11
    ||| :|_|:|
Db 70 TSHLRVHTGEK 80

RESULT 14
US-09-920-804-2
; Sequence 2, Application US/09920804
; Patent No. US20020054846A1
; GENERAL INFORMATION:
; APPLICANT: Seino, Susumu; JCR Pharmaceuticals Co., Ltd.
; TITLE OF INVENTION: Sodium Ion-Driven Chloride/Bi-Carbonate Exchanger
; FILE REFERENCE: GP44
; CURRENT APPLICATION NUMBER: US/09/920,804
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 2
; LENGTH: 1088
; TYPE: PRT
; ORGANISM: Mus musculus
; US-09-920-804-2

Query Match 54.7%; Score 35; DB 10; Length 1088;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
Qy 1 TSPLNIHNGQKL 12
    |||:||| :|
Db 620 TYPINMHNDLE 631

RESULT 15
US-10-010-901-29
; Sequence 29, Application US/10010901
; Patent No. US20020058201A1
; GENERAL INFORMATION:
; APPLICANT: McFadden, Grant
; TITLE OF INVENTION: NOVEL MYXOMA GENES FOR IMMUNE MODULATION
; FILE REFERENCE: 50082/009002
; CURRENT APPLICATION NUMBER: US/10/010,901
; CURRENT FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: US/09/615,041
; PRIOR FILING DATE: 2000-07-12
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FASTSEQ for Windows version 4.0
; SEQ ID NO 29
; LENGTH: 2000
; TYPE: PRT
; ORGANISM: Myxoma virus
; US-10-010-901-29

Query Match 54.7%; Score 35; DB 12; Length 2000;
Best Local Similarity 54.5%; Pred. No. 2.7e+02;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Qy 1 TSPLNIHNGQK 11
    |||:||| :|
Db 230 TLPVSIHQGSK 240

```

Search completed: January 3, 2003, 14:26:44
 Job time : 11 secs

GenCore version 5.1.3
 Copyright (c) 1993 - 2003 Compugen Ltd.

Om protein - protein search, using sw model

Run on: January 3, 2003, 14:23:57 ; Search time 29 seconds
 (without alignments)
 85.261 Million cell updates/sec

Title: US-09-899-376-1

Perfect score: 64

Sequence: 1 TSPLNINHGQKL 12

Scoring table: BLOSUM62
 Gapext 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : SPTREMBL_21;*

1: sp_archaea;*
 2: sp_bacteria;*
 3: sp_fungi;*
 4: sp_human;*
 5: sp_invertebrate;*
 6: sp_mammal;*
 7: sp_mhc;*
 8: sp_organelle;*
 9: sp_phage;*
 10: sp_plant;*
 11: sp_rabbit;*
 12: sp_virus;*
 13: sp_vertebrate;*
 14: sp_unclassified;*
 15: sp_rvirus;*
 16: sp_bacteriop;*
 17: sp_archeap;*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Query Match Length DB ID Description

1 43 67.2 321 4 Q96ER2 Q96ER2 homo sapien
 2 42 65.6 1343 16 Q8XAY4 Q8XAY4 escherichia
 3 42 65.6 2531 5 Q16004 O16004 lyttechninus
 4 41 64.1 427 10 Q9SZV7 Q9SZV7 arabidopsis
 5 40 62.5 331 3 Q14097 Q14097 schizosaccharomyces
 6 40 62.5 334 5 Q9GU28 Q9GU28 oikopleura
 7 39 60.9 243 16 Q8ZHG4 Q8ZHG4 yersinia pe
 8 39 60.9 392 4 Q9HB07 Q9HB07 homo sapien
 9 39 60.9 563 16 Q9X8S9 Q9X8S9 streptomyce
 10 39 60.9 738 4 Q9U1AL Q9U1AL homo sapien
 11 39 60.9 855 17 Q8T9P99 Q8T9P99 methanococcus
 12 39 60.9 914 12 Q9TF30 Q9TF30 bovine aden
 13 39 60.9 938 12 Q9W51 Q9W51 porcine ade
 14 39 60.9 939 12 Q8A178 Q8A178 porcine ade
 15 39 60.9 970 5 Q96115 Q96115 drosophila
 16 39 60.9 2176 5 Q9VGPI Q9VGPI drosophila

RESULT 2

8

17 38 59.4 147 11 Q9C016 Q9C016 mus musculus
 18 38 59.4 150 16 Q9C0B1 Q9C0B1 lactococcus
 19 38 59.4 173 4 Q9UKC5 Q9UKC5 homo sapien
 20 38 59.4 188 2 Q93MT1 Q93MT1 photorhabdu
 21 38 59.4 277 2 Q8VTA6 Q8VTA6 synechococcus
 22 38 59.4 324 5 Q22161 Q22161 caenorhabdi
 23 38 59.4 367 11 Q9V1V9 Q9V1V9 mus musculus
 24 38 59.4 376 4 Q96CK9 Q96CK9 homo sapien
 25 38 59.4 406 16 Q8ZMW4 Q8ZMW4 salmonella
 26 38 59.4 405 16 Q8ZAT6 Q8ZAT6 salmonella
 27 38 59.4 415 11 Q9JTE4 Q9JTE4 mus musculus
 28 38 59.4 444 10 P93716 P93716 petunia hybrida
 29 38 59.4 471 4 Q9H0V2 Q9H0V2 homo sapien
 30 38 59.4 480 11 Q9D663 Q9D663 mus musculus
 31 38 59.4 480 11 Q9D2J5 Q9D2J5 mus musculus
 32 38 59.4 641 5 Q18586 Q18586 caenorhabdi
 33 38 59.4 693 10 Q9SK16 Q9SK16 arabidopsis
 34 38 59.4 693 11 Q9X8E1 Q9X8E1 escherichia
 35 37 57.8 112 4 Q9H4Y3 Q9H4Y3 homo sapien
 36 37 57.8 177 5 Q9582 Q9582 mus musculus
 37 37 57.8 240 16 Q9694 Q9694 pseudomonas
 38 37 57.8 248 2 Q87957 Q87957 salmonella
 39 37 57.8 248 12 Q9J8C4 Q9J8C4 spodoptera
 40 37 57.8 299 2 Q9F7K7 Q9F7K7 salmonella
 41 37 57.8 299 2 Q9F7K4 Q9F7K4 salmonella
 42 37 57.8 299 2 Q9F7K0 Q9F7K0 salmonella
 43 37 57.8 299 2 Q9F7J7 Q9F7J7 salmonella
 44 37 57.8 299 2 Q9F719 Q9F719 salmonella
 45 37 57.8 299 2 Q9F714 Q9F714 salmonella

ALIGMENTS

RESULT 1

ID Q96ER2 PRELIMINARY; PRT; 321 AA.

AC Q96ER2; DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Unknown (Protein for MGC:21259)
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
 OC NCBI_TaxID=9606;
 RN [1] SEQUENCE FROM N.A.
 RC TISSUE=TESTIS, AND EMBRYONAL CARCINOMA;
 RA Strausberg R.; Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC012012; AAH12012; 1; -.
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR00822; Znf_C2H2.
 DR Pfam: PF01352; KRAB; 1.
 DR Pfam: PF00056; Znf_C2H2; 7.
 DR Pfam: PF00053; Znf_C2H2; 6.
 DR PROSITE: PS50055; KRAB; 1.
 DR PROSITE: PS50025; ZINC_FINGER_C2H2_1; UNKNOWN_6.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 7.
 KW DNA binding; zinc-finger.
 SEQUENCE 321 AA; 36369 MW; 62B8342C8B7067D9 CRC64;
 Query Match 67.2%; Score 43; DB 4; Length 321;
 Best Local Similarity 80.0%; Pred. No. 5.3; Mismatches 2; Indels 0; Gaps 0;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QBXY4	PRELIMINARY;	PRT; 1343 AA.
ID		DR
QBXY4		InterPro; IPR00561; EGF-like.
AC		DR
DT	01-MAR-2002 (TREMBlrel. 20, Created)	InterPro; IPR00742; EGF-2.
DT	01-MAR-2002 (TREMBlrel. 20, Last sequence update)	InterPro; IPR01881; EGF-Ca.
DT	01-MAR-2002 (TREMBlrel. 20, Last annotation update)	InterPro; IPR01438; EGF-II.
DE	Orf, hypothetical protein.	InterPro; IPR00800; Notch.
DR	YDK OR Z2195 OR EGS217.	Pfam; PF00023; ankyrin; 6.
GN	Escherichia coli O157:H7.	DR
OS	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;	Pfam; PF00066; notch; 3.
OC	Escherichia.	PRINTS; PRO0415; ANKYRIN.
OX	NCBI_TaxID=33334;	PRINTS; PRO0010; EGFBLLOOD.
RN		PRINTS; PRO1452; NOTCH.
RP		SMART; SM00248; ANK; 5.
RC		SMART; SM00179; EGF-CA; 23.
RX		SMART; SM0019; EGF-CA; 23.
RA		SMART; SM0004; NL; 3.
RA		PROSITE; PS00088; ANK_REPEAT; 10.
Rose D. J., Mayhew G. F., Evans P. S., Gregor J. J., Kirkpatrick H. A.,		DR
Rose D. J., Mayhew G. F., Evans P. S., Gregor J. J., Kirkpatrick H. A.,		PROSITE; PS00297; ANK REP REGION; 2.
Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,		DR
Grotbeck E. J., Davis N. W., Lim A., Bimalanta E. T., Potamouidis K.,		PROSITE; PS00010; ASX_HYDROXYL; 21.
R.A. Apodaca J., Anantharaman T. S., Lin J., Yen G., Schwartz D.C.,		DR
R.A. Welch R. A., Blattner F. R.,		PROSITE; PS00022; EGF-1; UNKNOWN_33.
R.T. "Genome sequence of enterohemorrhagic Escherichia coli O157:H7,"		DR
RL	Nature 409:529-533 (2001).	PROSITE; PS01187; EGF-CA; 25.
RL		ANR repeat; calcium-binding; EGF-like domain; Glycoprotein;
RN	[2]	Hydroxylation; Repeat; 2331 AA; 273982 MW; 5BF42BEC627CA303 CRC64;
RP		SEQUENCE FROM N.A.
RC		SPRAIN-0157:H7 / RIMD 0509952;
RX		MEDLINE=21156231; PubMed=11258796;
RA		Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA		Han C.-G., Ohtsuka E., Nakaya K., Murata T., Tanaka M., Tobe T.,
R.A. Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,		DR
R.A. Kohara S., Shiba T., Hattori M., Shinagawa H.,		PROSITE; PS00116; EGF-2; 25.
RT	"Complete genome sequence of enterohemorrhagic Escherichia coli	DR
RT	O157:H7 and genomic comparison with a laboratory strain K-12."	PROSITE; PS01187; EGF-CA; 25.
RL	DNA Res 8:11-22 (2001).	ANR repeat; calcium-binding; EGF-like domain; Glycoprotein;
DR	EMBL; AE005354; AAG56256.1; -.	Hydroxylation; Repeat; 2331 AA; 273982 MW; 5BF42BEC627CA303 CRC64;
KW	Complete proteome	SEQUENCE FROM N.A.
SQ	SEQUENCE - 1343 AA; 138568 MW; 4C2456739907B63D CRC64;	SPRAIN-0157:H7 / RIMD 0509952;
Query	Match	SPRAIN-0157:H7 / RIMD 0509952;
Best	Local Similarity	SPRAIN-0157:H7 / RIMD 0509952;
Matches	88.9%; Pred. No. 35;	SPRAIN-0157:H7 / RIMD 0509952;
QY	8; Conservative	SPRAIN-0157:H7 / RIMD 0509952;
Db	0; Mismatches	SPRAIN-0157:H7 / RIMD 0509952;
4	1; Indels	SPRAIN-0157:H7 / RIMD 0509952;
Db	0; Gaps	SPRAIN-0157:H7 / RIMD 0509952;
991	0; Oligo	SPRAIN-0157:H7 / RIMD 0509952;
LNIHNGOKL	999	SPRAIN-0157:H7 / RIMD 0509952;
RESULT 3		RESULT 4
O16004	PRELIMINARY;	PRELIMINARY;
ID	PRT; 2531 AA.	PRT; 427 AA.
AC		DR
AC		Q9S2V7;
DT	01-JAN-1998 (TREMBlrel. 05, Created)	DR
DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)	Q9S2V7;
DE	Notch homolog.	01-MAY-2000 (TREMBlrel. 13, Last sequence update)
OS	Lytrochinius variegatus (Sea urchin).	01-MAY-2000 (TREMBlrel. 13, Last annotation update)
OC	Bukaryota; Metacosa; Echinodermata; Eleutherozoa; Echinozoa;	01-JUN-2002 (TREMBlrel. 21, Last annotation update)
OC	Echinoidea; Euechinoidea; Echinacea; Temnopleuroidea; Toxopneustidae;	Hypothetical 47.4 kDa protein.
OC	Lytrochinius.	GN F6G3.70 OR AT4G30040.
OC	NCBI_TaxID=7654;	OS Arabidopsis thaliana (Mouse-ear cress).
RN		OC Eukaryota; Viridiplantae; Streptophyta; eudicots; core eudicots; Rosidae;
RN		OC Spermatophyta; Magnoliophyta; eudicots; core eudicots; Rosidae;
RN		OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
RN		OX NCBL_TaxID=3702;
RN		RN
SEQUENCE FROM N.A.		SEQUENCE FROM N.A.
RA	Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,	RA
RA	Mayer K.F.X., Lemcke K., Schueler C.,	RA
RA	Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.	RA
RN		RN
RN		RN
SEQUENCE FROM N.A.		SEQUENCE FROM N.A.
RA	EU Arabidopsis	RA
RA	Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.	RA
RN		RN
SEQUENCE FROM N.A.		SEQUENCE FROM N.A.
RA	Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.,	RA
RA	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.	RA
RN		RN
SEQUENCE FROM N.A.		SEQUENCE FROM N.A.
RA	EU Arabidopsis sequencing project;	RA
RA	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.	RA
RN		RN
SEQUENCE FROM N.A.		SEQUENCE FROM N.A.
RA	Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.,	RA
RA	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.	RA
RN		RN
SEQUENCE FROM N.A.		SEQUENCE FROM N.A.
RA	EU Arabidopsis sequencing project;	RA
RA	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.	RA
RN		RN
SEQUENCE FROM N.A.		SEQUENCE FROM N.A.
RA	DR EMBL; AL078464; CAB43839.1; -.	RA
RA	DR EMBL; AL61576; CAB8097.1; -.	DR
RN		DR InterPro; IPR001461; AspproteaseAI.
RT	RT insights into vegetal plate regionalization and Notch receptor regulation.	DR Pfam; PF00026; asp; 1.
RT	RT development.	DR PRINTS; PR00792; PEPSIN.
RL	RL HSSP; P01132; IEGF.	KW Hypothetical protein.
DR	DR InterPro; IPR002110; ANK.	SQ SEQUENCE 427 AA; 47357 MW; F6F42BDD938B3225 CRC64;

Qy	1	TSPLNIHNG 9	64.1%; Score 41; DB 10; Length 427; Best Local Similarity 77.8%; Pred. No. 17; Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db	263	TTPLIEHNG 271	
RESULT 5			
Q4097		PRELIMINARY; PRT; 331 AA.	
ID	014097;		
AC	Q14097;		
DT	01-NOV-1998 (TREMBlrel. 08, Last sequence update)		
DT	01-MAR-2002 (TREMBlrel. 20, Last annotation update)		
DE	Hypothetical protein C2F3_14C in chromosome I.		
GN	SPAC2F3_14C.		
OS	Schizosaccharomyces pombe (Fission yeast).		
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetidae;		
OC	Schizosaccharomycetidae; Schizosaccharomyces.		
OX	NCBI_TAXID=4896;		
RN			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=972;		
RA	Oliver K., Harris D., Barrell B.G., Rajandream M.A., Wood V., Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.		
RL	Z99165; CAB16268.1; -; InterPro: IPR001202; WW_Rsp5_WWP.		
DR	Protein: PR0097; WW; 1.		
KW	Hypothetical protein.		
SEQUENCE	331 AA; 37669 MW; EA888EACC4CBB7B1 CRC64;		
Query Match	62.5%; Score 40; DB 3; Length 331; Best Local Similarity 72.7%; Pred. No. 20; Matches 8; Conservative 0; Mismatches 0; Indels 3; Gaps 0;		
Qy	2	SPLNTHNGKL 12	
Db	271 SPLASHNGSL 281		
RESULT 6			
Q9GU28		PRELIMINARY; PRT; 334 AA.	
ID	09GU28		
AC	Q9GU28;		
DT	01-MAR-2001 (TREMBlrel. 16, Created)		
DT	01-MAR-2001 (TREMBlrel. 16, Last sequence update)		
DT	01-MAR-2002 (TREMBlrel. 20, Last annotation update)		
DE	Brachyury protein.		
GN	Q9GU28		
OS	Brachyury protein.		
OC	Brachyury (m) expression in embryos of a larvacean urochordate, Oikopleura dioica, and the ancestral role of brachyury.;		
OC	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.		
OX	NCBI_TAXID=34765;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Basham S., Postlethwait J.H.;		
RT	Oikopleura dioica; Chordata; Urochordata; Appendicularia; Oikopleura; Oikopleura.		
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.		
DR	AF204208; AAG22592.1; -.		
RSP	P24781; IXSR;		
DR	InterPro: IPR001699; TF_T-box.		
DR	Prints; PR00907; T-Box; 1.		
DR	SMART; SM00425; TBOX; 1.		
DR	PROSITE; PS01283; TBOX_1; 1.		
DR	PS0252; TBOX_3; 1.		
SEQUENCE	334 AA; 37574 MW; 02AEA233B5800F50 CRC64;		
Query Match	62.5%; Score 40; DB 5; Length 334; Best Local Similarity 58.3%; Pred. No. 21; Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;		
Qy	1	TSPLNIHNGQKL 12	
Db	125 TNKLNQHNGQQI 136		
RESULT 7			
Q8ZH64		PRELIMINARY; PRT; 243 AA.	
ID	08ZH64		
AC	Q8ZH64;		
DT	01-MAR-2002 (TREMBlrel. 20, Created)		
DT	01-MAR-2002 (TREMBlrel. 20, Last sequence update)		
DE	Hypothetical protein YP00934.		
GN	YPO0934.		
OS	Yersinia pestis.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Yersinia.		
OC	NCBI_TAXID=622;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CO-92 / BIOVAR ORIENTALIS;		
RC	MEDLINE=2147013; PubMed=11586360;		
RA	Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G., Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L., Baker S., Basham D., Bentley D., Brooks D., Cerdeno-Tarraga A.M., Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G., Feltwell T., Hamlin N., Holroyd S., Jigels K., Karlyshev A.V., Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;		
RA	Nature 413:523-527(2001).		
RA	EMBL; AJ414143; CAD89777.1; -.		
DR	InterPro: IPR004382; Cons_hypo46; 1.		
DR	TIGRFAMS; TIGR00046; Cons_hypo46; 1.		
KW	Hypothetical protein; Complete proteome.		
SEQUENCE	243 AA; 26920 MW; BF54ET09F9A37CC CRC64;		
Query Match	60.9%; Score 39; DB 16; Length 243; Best Local Similarity 77.8%; Pred. No. 23; Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;		
Qy	2	SPLNTHNGO 10	
Db	76 SPLNUHLGQ 84		
RESULT 8			
Q9HB07		PRELIMINARY; PRT; 392 AA.	
ID	Q9HB07		
AC	Q9HB07;		
DT	01-MAR-2001 (TREMBlrel. 16, Created)		
DT	01-MAR-2001 (TREMBlrel. 16, Last sequence update)		
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)		
DE	CDNA FLJ14011 fis, clone Y79AA1002472, weakly similar to zinc finger protein 91.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Butteleostomi; Mammalia; Eutheria; Primates; Catarhini; Hominidae; Homo.		
OC	NCBI_TAXID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Isocai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Wagatsuma M., Hssoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M., Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S., Watamabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Niinomiya K., Iwayanagi T.;		
RA	"NEDO human cDNA sequencing project.;"		
RT			

RL	Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
CC	-; SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).	OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
-;		OX	NCBI_TAXID=9606;
EMBL	AK024073; BAB14816.1; -.	RN	[1]
HSSP	P08046; 1A1H.	RP	SEQUENCE FROM N.A.
InterPro	IPR000822; Znf_C2H2; 13.	RA	Lamerdin J.E., McCreedy P.M., Skowronski E., Viswanathan V.,
PRNTS	PRNTS: PRO0048; ZINC-FINGER	RA	Pham; P000003; Znf_C2H2; 7.
PRDM	PRDM; P000003; Znf_C2H2; 7.	RA	Burkhardt-Schulte K., Gordon L., Dias J., Ramirez M., Stilwagen S.,
SMART	SMART; SM00355; Znf_C2H2; 12.	RA	Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
PROSITE	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 10.	RA	Danganan L., Erlle A., Christensen M., Georgescu A., Avila J., Liu S.,
PROSITE	PROSITE; PS0157; ZINC_FINGER_C2H2_2; 12.	RA	Attix C., Andreise T., Frankheim M., Amico-Keller G., Coiffield J.,
DR	DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.	RA	Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
KW	DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.	RA	Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
SQ	SEQUENCE 392 AA; 45453 MW; 3398446D50082D2D CRC64;	RA	Olsen A.S., Carrano A.V.,
Query	Query Match 60.9%; Score 39; DB 4; Length 392;	RT	"Sequence analysis of a 2.3 Mb region containing a zinc finger (ZNF)
Best	Best Local Similarity 77.8%; Pred. No. 37;	RT	cluster in 19q13.4.";
Matches	Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;	RL	Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
QY	4 LNTHNGQKL 12	CC	! SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
Db	159 LRTIHNGEKL 167	EMBL	AC013256; RAAF06067; 1; -.
RESULT	RESULT 9	DR	HSSP; P08046; 1A1J.
RP	09X8S9	DR	InterPro; IPR01909; KRAB.
RC	09X8S9	DR	InterPro; IPR00822; Znf_C2H2.
AC	PRELIMINARY;	DR	Pfam; PR01352; KRAB; 1.
DT	01-NOV-1999 (TREMBLREL. 12, Created)	DR	Pfam; PR00096; zf-C2H2; 16.
DT	01-JUN-1999 (TREMBLREL. 12, Last sequence update)	DR	PRNTS; PR00048; ZINC-FINGER.
DT	01-JUN-2002 (TREMBLREL. 21, Last annotation update)	DR	PRNTS: PD000003; Znf_C2H2; 7.
DE	Hypothetical protein SCO3893.	DR	SMART; SM00349; KRAB; 1.
GN	SCO3893 OR SCH24_15C.	DR	SMART; SM00355; Znf_C2H2; 15.
OS	Streptomyces coelicolor.	DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 14.
OC	Bacteria; Firmicutes; Actinobacteria; Actinomycetidae; Streptomycetes.	DR	PROSITE; PS0157; ZINC_FINGER_C2H2_2; 15.
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.	DR	DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
OX	[1]	SQ	SEQUENCE 738 AA; 84252 MW; 9E158059D5C18527 CRC64;
RN	SEQUENCE FROM N.A.	Query	Query Match 60.9%; Score 39; DB 4; Length 738;
RP	STRAINFA3(2) / M145;	Best	Best Local Similarity 77.8%; Pred. No. 71;
RC	Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,	Matches	Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
RA	Thompson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,	QY	4 LNTHNGQKL 12
RA	Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,	Db	505 LRTIHNGEKL 513
RA	Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,	RESULT	RESULT 11
RA	Huanc C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,	RP	08T99
RA	Rabinowitzsch E., Rajandream M.A., Rutherford K., Rutter S.,	AC	08T99;
RA	Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,	08T99;	PRELIMINARY;
RA	Warren T., Wetzorek A., Woodward J., Barrell B.G., Parkhill J.,	08T99;	FRT;
RA	Hopwood D.A.;	08T99;	855 AA.
RT	Complete genome sequence of the model actinomycete Streptomyces	DT	01-JUN-2002 (TREMBLREL. 21, Created)
RT	ceticolor A3(2)";	DT	01-JUN-2002 (TREMBLREL. 21, Last sequence update)
RT	Nature 417:141-147 (2002).	DE	Peptidase family protein U32.
RL	EMBL; AL049826; CAB42719.1; -.	GN	MA0538.
DR	Hypothetical protein.	OS	Methanobarcina acetylavorans.
SQ	SEQUENCE 563 AA; 61226 MW; 7FB4DDF8B5F4EB72 CRC64;	OC	Archaea; Euryarchaeota; Methanococci; Methanomicrobiales;
Query	Match 60.9%; Score 39; DB 16; Length 563;	OC	Methanomicrobiae; Methanomarcina.
Best	Local Similarity 58.3%; Pred. No. 54;	OX	NCBI_TaxID=2214;
Matches	Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;	RN	[1]
QY	1 TSPLNTHNGQKL 12	RP	SEQUENCE FROM N.A.
Db	57 TSPPELHSGHKL 68	RC	STRAINFC2A / ATCC 35395 / DSM 2834;
RESULT	RESULT 10	RX	MEDLINE-21929760; PubMed=11932288;
Q9ULAA	Q9ULAA PRELIMINARY;	RA	Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., McDonald P.,
ID	Q9ULAA	RA	Fitzhugh W., Calvo S., Engels F., Smirnov S., Attnor D., Johnson R.,
AC	Q9ULAA	RA	Allen N., Nevalo J., Strange-Thomann N., DeArleano K., Johnson R.,
DT	01-MAY-2000 (TREMBLREL. 13, Created)	RA	Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
DT	01-MAY-2000 (TREMBLREL. 13, Last sequence update)	RA	Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.E., Guss A.M.,
DT	01-JUN-2002 (TREMBLREL. 21, Last annotation update)	RA	Heidecker R., Ingram-Smith C., Kuethe H.C., Krzycki J.A., Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
DE	R31155_1.	RA	Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C., Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I., Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E., Metcalf W.W., Birren B.;
OS	Homo sapiens (Human).	RT	"The genome of Methanobarcina acetylavorans reveals extensive metabolic and physiological diversity";
RL		RT	Genome Res. 12:532-542(2002).

DR	EMBL: AE010714; AAM03982.1; -.	DR	Pfam; PF01065; Adeno_hexon; 1.
KW	complete proteome.	DR	Pfam; PF03678; Adeno_hexon_C; 1.
SQ	SEQUENCE 855 AA; 94070 MW; 19CFE36AB92AC075 CRC64;	DR	prodom; PD002815; Adeno_hexon; 1.
Query Match	60.9%; Score 39; DB 12; Length 855;	SQ	SEQUENCE 938 AA; 105964 MW; 8115164CD481DDBA CRC64;
Best Local Similarity	87.5%; Pred. No. 83; Mismatches 0; Indels 0; Gaps 0;	Query Match	60.9%; Score 39; DB 12; Length 938;
Matches	7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;	Best Local Similarity 77.8%; Pred. No. 91; Mismatches 2; Indels 0; Gaps 0;	
Qy	2 SPLNTHNG 9	Qy	2 SPLNTHNGQ 10
Db	676 SPLNLYNG 683	Db	223 SPTNTHGQ 231
RESULT 12			
091F30	PRELIMINARY; PRT; 914 AA.	094178	PRELIMINARY; PRT; 939 AA.
ID	091F30; 01-OCT-2000 (TREMBLrel. 15, Created)	ID	094178; 01-NOV-1996 (TREMBLrel. 01, Created)
AC	091F30; (TREMBLrel. 15, last sequence update)	AC	094178; 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT	01-OCT-2000 (TREMBLrel. 15, last sequence update)	DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)	DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE	Hxen protein (Fragment).	DE	Hxen.
OS	Bovine adenovirus type 10 (Mastadenovirus bos10).	OS	porcine adenovirus type 3 (PAV-3).
OC	Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.	OC	Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX	NCBI_TaxID=3988;	OX	NCBI_TaxID=35265;
RN	[1]	RN	[1]
RP	SEQUENCE FROM N.A.	RP	SEQUENCE FROM N.A.
RA	Lehmkuhl H.D., Hobbs L.A.;	RA	McCoy R.J., Johnson M.A., Sheppard M.;
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.	RL	Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF242774; AAF21361; -.	DR	EMBL; U34592; AAB02183; 1; -.
DR	HSSP: P03277; 1DHX.	DR	HSSP: P03277; 1DHX.
DR	InterPro; IPR000736; Adeno_hexon.	DR	InterPro; IPR00736; Adeno_hexon.
DR	Pfam; PF01065; Adeno_hexon; 1.	DR	Pfam; PF01065; Adeno_hexon; 1.
DR	ProDom; PF03678; Adeno_hexon_C; 1.	DR	ProDom; PF03678; Adeno_hexon_C; 1.
DR	ProDom; PF002815; Adeno_hexon; 1.	DR	ProDom; PD002815; Adeno_hexon; 1.
FT	NON_TER 914	FT	SEQUENCE 939 AA; 105897 MW; 10DD0854972DC099 CRC64;
SQ	SEQUENCE 914 AA; 103905 MW; 5508E006997739CD CRC64;	SQ	SEQUENCE 939 AA; 105897 MW; 10DD0854972DC099 CRC64;
Query Match			
Best Local Similarity	60.9%; Score 39; DB 12; Length 914;	Best Local Similarity	60.9%; Score 39; DB 12; Length 939;
Matches	77.8%; Pred. No. 89; Mismatches 2; Indels 0; Gaps 0;	Matches	77.8%; Pred. No. 91; Mismatches 2; Indels 0; Gaps 0;
Qy	2 SPLNTHNGQ 10	Qy	2 SPLNTHNGQ 10
Db	217 SPTNTHGQ 225	Db	223 SPTNTHGQ 231
RESULT 13			
09WB51	PRELIMINARY; PRT; 938 AA.	096115	PRELIMINARY; PRT; 970 AA.
ID	09WB51; 01-NOV-1999 (TREMBLrel. 12, Created)	ID	096115; 01-DEC-2001 (TREMBLrel. 19, Created)
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)	DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)	DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE	REXON.	DE	GH21817P.
OS	Porcine adenovirus type 3 (PAV-3).	OS	CG270.
OC	Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.	OC	drosophila melanogaster (Fruit fly).
OX	NCBI_TaxID=35265;	OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydriidae; Diopsophilidae; Drosophila.
RP	SEQUENCE FROM N.A.	OX	NCBI_TaxID=7227;
RC	STRAIN=IAF;	RN	[1]
RA	Larocque D., Malenfant F., Massie B., Dea S.;	RP	SEQUENCE FROM N.A.
RT	"porcine adenovirus serotype 3, complete genome.";	RC	STRAIN=BERKELEY;
RL	Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.	RA	Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J., Champe M., Chavez C., Dorrest V., Farfan D., Frise E., George R., Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mundall C.J., Nunoo J., Pacieb J., Pararas V., Park S., Phouanenavong S., Wan K., Yu C., Lewis M.E., Rubin G.M., Celniker S., Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RR	[2]	RA	EMBL; AJ231815; CAB40301; -.
RA	Larocque D.;	RA	EMBL; AB051617; BAA76968; 1; -.
RT	"porcine adenovirus serotype 3, complete genome.";	RA	FlyBase; FBgn0037897; CG9941; -.
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.	DR	FlyBase; FBgn0037897; CG570.
DR	EMBL; AJ231815; CAB40301; -.	DR	InterPro; IPR000306; Znf_FIVE.
DR	HSSP; P03277; 1DHX.	DR	Pfam; PF01363; FYVE; 1.
DR	InterPro; IPR000736; Adeno_hexon.		

Fri Jan 3 14:42:50 2003

us-09-899-376-1.rspt

Page 6

SQ SEQUENCE 970 AA: 108912 MW: 7FACD0DE9852607 CRC64:
Query Match 60.9%; Score 39; DB 5; Length 970;
Best Local Similarity 66.7%; Pred. No. 94;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Qy 1 TSPTNTHNGOKL 12
| || | | :||
Db 374 TGPPLNNHNLRLK 385

Search completed: January 3, 2003, 14:26:07
Job time: 31 secs

GenCore version 5.1.3									
copyright (c) 1993 - 2003 Compugen Ltd.									
OM protein - protein search, using sw model									
Run on: January 3, 2003, 14:22:12 ; search time 11 seconds									
Total number of hits satisfying chosen parameters:	112892	Score:	64	US-09-899-376-1	Sequence:	1 TSPLNTHNGOKL 12	Scoring table:	BLOSSUM62	Scoring table:
Minimum DB seq length:	0	Gapop 10.0 , Gapext 0.5			Searched:		Minimum DB seq length:	0	Maximum DB seq length: 200000000
Post-processing:	Minimum Match 0%	Maximum Match 100%	Post-processing:	Minimum Match 0%	Maximum Match 100%	Post-processing:	DE	DE	DE
Database :	Swissprot_40.*	Listing first 45 summaries	Database :	Swissprot_40.*	Listing first 45 summaries	Database :	rhamnose synthetase	rhamnose synthetase	rhamnose synthetase
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
SUMMARIES									
Result No.	Score	Query Match	Length	DB ID	Description	Result 1	ID	RFBD_SHFL	STANDARD;
1	42	65.6	300	1 RFBBD_SHFL	P37778 shigella fl	P37778:	AC	P37778;	
2	42	65.6	1302	1 MDR5_DROME	Q00748 drosophila	DT 01-OCT-1994 (Rel. 30, Created)	DT		
3	42	65.6	1325	1 YDEK_ECOLI	P32051 escherichia	DT 01-OCT-1994 (Rel. 30, Last sequence update)	DT		
4	39	60.9	939	1 HEK_ADPBP3	Q9YTR8 porcine ade	DE dtDP-4-dehydrorhamnose reductase (EC 1.1.1.133) (dtDP-4-keto-L-rhamnose reductase) (dtDP-6-deoxy-L-mannose dehydrogenase) (dtDP-L-rhamnose synthetase).	DE		
5	38	59.4	223	1 GLT4_HUMAN	Q13360 homo sapien	DE RFBP.	DE		
6	38	59.4	321	1 Z177_HUMAN	Q9uk99 homo sapien	GN RFBP.	GN		
7	38	59.4	471	1 FBX3_HUMAN	P1075 homo sapien	OS Bacteria: Shigella flexneri.	OS		
8	38	59.4	683	1 PLO1_SC1PHO	P50528 schizosaccharomyces pombe	OC Shigella: Proteobacteria; gamma subdivision; Enterobacteriaceae;	OC		
9	37	57.8	111	1 SOHA_ECOLI	P15373 escherichia	OC Shigella. NCBI_TaxID=623;	OC		
10	37	57.8	299	1 RFBBD_SA0TY	P26392 salmonella	RN [1]	RN		
11	37	57.8	476	1 GAG_B127	P19558 bovine immunoglobulin G	RP SEQUENCE FROM N.A.	RP		
12	37	57.8	476	1 GAG_B127	P19559 bovine immunoglobulin G	RC STRAIN-Serotype 2A;	RC		
13	36	194	1 MINC_HMPJP	Q9zm51 helicobacte	CC	RX MEDLINE:9424146; PubMed=8170390;	RX		
14	36	56.2	195	1 MINC_HMPJP	Q25693 helicobacte	RA Macpherson D.F., Manning P.A., Morona R.; "Characterization of the dtDP-rhamnose biosynthetic genes encoded in the rfb locus of shigella flexneri." Mol. Microbiol. 11:281-292(1994).	RA		
15	36	56.2	243	1 YGGJ_ECOLI	P57912 escherichia	CC -1- CATALYTIC ACTIVITY: dtDP-6-deoxy-L-mannose + NADP(+)= dtDP-4-dehydro-6-deoxy-L-mannose + NADP.	CC		
16	36	56.2	253	1 Y110_B1C1AT	P57488 buchnera ap	CC -!- COFACTOR: NADP.	CC		
17	36	56.2	307	1 DNAJ_SYN3	P73097 synchocystis	CC BIOSYNTHESIS PATHWAY OF LIPOPOLYSACCHARIDE BIOSYNTHESIS.	CC		
18	36	56.2	430	1 SUN_COBEU	P45679 coxiella bu	CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation in the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to licensed@isb-sib.ch).	CC		
19	36	60.5	1 Y598_TREPA	Q83607 treponema p	CC	DR EMBL: X71970; CAA50768.1; -.	DR		
20	36	60.5	1005	1 BGAL_ACTPL	P70753 actinobacil	DR PIR: S41535; S41555.	DR		
21	35	54.7	288	1 DDM4A_MINTSP	P21161 metylobacil	DR TIGRFAMS; TIGR01214; rmlD; 1.	DR		
22	35	54.7	455	1 HN4A_XENLA	P091766 xenopus lae	DR Li-polysaccharide biosynthesis; Oxidoreductase; NADP.	DR		
23	35	54.7	491	1 G6PD_BUCA1	P57405 buchnera ap	FT NP_BIND 7	FT		
24	35	54.7	551	1 SVE_ARCFU	P029979 archaeboglob	FT NP_BIND 7	FT		
25	35	54.7	928	1 YDG1_SC1PHO	Q010491 schizosaccharomyces pombe	FT SEQUENCE 300 AA: 32757 MW: 0F9ABB16AF83AF99 CRC64;	FT		
26	35	54.7	942	1 DPOL1_CHL40	Q008307 chloroflexu	QY 1 TSPLNTHNGOKL 12	QY		
27	34	53.1	93	1 VGGJ_ECOLI	P37795 erwinia chri	Db 122 TAPLNVYGGTKL 133	Db		
28	34	53.1	205	1 LEXA_PBORE	P007267 providencia	RESULT 2	RESULT 2		
29	34	53.1	223	1 GDX1_TOBAC	P03362 nicotiana t	ID MDR5_DROME STANDARD; PRT: 1302 AA.	ID MDR5_DROME STANDARD; PRT: 1302 AA.		
30	34	53.1	285	1 YFO_YEAST	P040186 saccharomyces cerevisiae				
31	34	53.1	287	1 KDSA_CAUCL	P37760 escherichia				
32	34	53.1	299	1 RBD1_ECOLI	P046769 escherichia				
33	34	53.1	301	1 RBD2_ECOLI	P09jw50 neisseria m				

AC	000748: 097FW3;
DT	01-APR-1993 (rel. 25, Created)
DT	15-JUN-2002 (rel. 41, Last sequence update)
DE	Multidrug resistance protein homolog 65 (P-glycoprotein 65).
OS	<i>Drosophila melanogaster</i> (Fruit fly)
OC	Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC	Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC	Muscomorpha; Ephydrioidea; Drosophilidae; Drosophila.
OK	NCBI_TAXID:7227;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE-Head;
RX	Wu C.-T., Budding M., Griffin M.S., Croop J.M.,
RX	Medline=9104385; PubMed=2072301;
RA	Begun D.J., Whitley P.;
RT	"Genetics of alpha-amanitin resistance in a natural population of <i>Drosophila melanogaster</i> ";
RL	Heredity 85:184-190(2000).
RN	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Berkeley;
RX	Medline=20196006; PubMed=10731132;
RA	Adams M.D., Celikiner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA	Ananatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA	Brannon R.C., Rogers Y.-H.C., Blazquez R.G., Champe M., Pfeiffer B.D.,
RA	Wen K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA	Abil J.F., Agbayani A., An H.-J., Andrews-Pflanck C., Baldwin D.,
RA	Bailey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA	Beeson K.Y., Beroz P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA	Burtis K.C., Busam D.A., Butler H., Cadile E., Center A., Chandra I.,
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA	de Pablo J., Delcher A., Deng L., Mays A.D., Dew I., Dietz S.M.,
RA	Dodson K., Doup L.E., Donnen M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA	Durbin R.K.J., Evangelista C.C., Ferrat C., Ferreria S., Fleischmann W.,
RA	Fosler C.J., Gabrilian A.E., Gang N.S., Gelbart W.M., Glasser K.,
RA	Gloed A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA	Houston D.M., Houston K.A., Howland T.J., Wei M.-H., Ibeawuchi C.,
RA	Jalil M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA	Lasko P., Lei Y., Levitt A.A., Li J., Li Z., Liang Y., Lin X.,
RA	Liu X., Marte B., McIntosh T.C., McLeod M.P., McPherson D.,
RA	Merkulov G., Mishina N.V., Mobarry C., Morris J., Moskrafi A.,
RA	Mouton S.M., Moy M., Murphy B., Murphy L., Muny D.M., Nelson D.L.,
RA	Nelson D.R., Nelson K.A., Nixon B.J., Nusskern D.R., Paciel J.M.,
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA	Shieh B.C., Siden-Kiamos I., Simpson R.M., Skupski M.P., Smith T.,
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA	Svartkars R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA	Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach B.J.,
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA	Zheng X.H., Zhong F.N., Zong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT	The genome sequence of <i>Drosophila melanogaster</i> .;
RL	Science 287:2185-2195 (2000).
CC	! SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC	! SURCELLULAR LOCATION: Integral membrane protein.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration

RESULT 3		CC	
YDEK_ECOLI		DR	DR
ID	YDEK_ECOLI	STANDARD;	FRT;
AC	P32451; P76140; P77168;	1325 AA.	
DT	01-NOV-1993 (Rel. 27, created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	16-APR-2001 (Rel. 40, Last annotation update)		
CC		between the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is modified and this statement is not removed. Usage by and entities requires a license agreement (See http://www.isb-sib.ch).	
CC		or send an email to license@ib-sib.ch).	
CC		-----	
DR	EMBL; M5077; AA28680; 1; -.	DR	DR
DR	EMBL; AF251287; AAF91471; -.	DR	DR
DR	EMBL; AF251285; AAF691461; -.	DR	DR
DR	EMBL; AE003563; AAF506691; -.	DR	DR
DR	PIR; B41249; B41249.	DR	DR
DR	FLYBASE; FBgn004513; MD-65.	DR	DR
DR	InterPro; IPR003593; ABC-ATPase.	DR	DR
DR	InterPro; IPR003439; ABC transporter.	DR	DR
DR	InterPro; IPR001140; ABC transporter.	DR	DR
DR	Pfam; PF00005; ABC_tran; 2.	DR	DR
DR	Pfam; PF00664; ABC_membrane; 2.	DR	DR
DR	ProDom; PD00006; ABC_transporter; 2.	DR	DR
DR	SMART; SM00382; AAA; 2.	DR	DR
DR	ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat; Multigene family.	DR	DR
FT	DOMAIN 1 48	CYTOPLASMIC (POTENTIAL).	
FT	TRANSMEM 49 69	1 (POTENTIAL).	
FT	DOMAIN 70 118	EXTRACELLULAR (POTENTIAL).	
FT	TRANSMEM 119 147	2 (POTENTIAL).	
FT	DOMAIN 148 194	CYTOPLASMIC (POTENTIAL).	
FT	TRANSMEM 195 215	3 (POTENTIAL).	
FT	DOMAIN 216 223	EXTRACELLULAR (POTENTIAL).	
FT	TRANSMEM 224 242	4 (POTENTIAL).	
FT	DOMAIN 243 302	CYTOPLASMIC (POTENTIAL).	
FT	TRANSMEM 303 323	5 (POTENTIAL).	
FT	DOMAIN 324 341	EXTRACELLULAR (POTENTIAL).	
FT	TRANSMEM 342 362	6 (POTENTIAL).	
FT	DOMAIN 363 731	CYTOPLASMIC (POTENTIAL).	
FT	TRANSMEM 732 753	7 (POTENTIAL).	
FT	DOMAIN 754 776	EXTRACELLULAR (POTENTIAL).	
FT	TRANSMEM 777 798	8 (POTENTIAL).	
FT	DOMAIN 799 852	CYTOPLASMIC (POTENTIAL).	
FT	TRANSMEM 853 873	9 (POTENTIAL).	
FT	DOMAIN 874 874	EXTRACELLULAR (POTENTIAL).	
FT	TRANSMEM 875 894	10 (POTENTIAL).	
FT	DOMAIN 895 956	CYTOPLASMIC (POTENTIAL).	
FT	TRANSMEM 957 977	11 (POTENTIAL).	
FT	DOMAIN 978 993	EXTRACELLULAR (POTENTIAL).	
FT	TRANSMEM 994 1014	12 (POTENTIAL).	
FT	DOMAIN 1015 1302	CYTOPLASMIC (POTENTIAL).	
FT	NP_BIND 440 447	ATP (POTENTIAL).	
FT	NP_BIND 1094 1101	ATP (POTENTIAL).	
FT	REPAT 671 673	1.	
FT	REPAT 674 1302	2.	
FT	CARBOHYD 103 103	N-LINKED (GLCNAC . .) (POTENTIAL).	
FT	CONFLICT 359 369	T -> S (IN REF. 1).	
FT	CONFLICT 678 678	F -> L (IN REF. 1).	
SO	SEQUENCE 1302 AA; 143784 MW;	3947-BBCBFA31924A CRC64;	
Query	Match 65, 68;	Score 42; DB 1; Length 1302;	
Best Local Similarity 72.7%;	Pred. No. 8, 5;		
Matches 8;	Conservative 0;	Mismatches 3;	Indels
QY	1 TSPNTHNOK 11		
QY	1 1111111111		
Db	673 TSPNFEKGOK 683		

RA Ruppert J.-M., Kinzler K.W., Wong A.J., Bigner S.H., Kao F.T.,
 RA Law M.L., Seuanez H.N., O'Brien S.J., Vogelstein B.;
 RT "The GLI-Kruppel family of human genes.";
 RL Mol. Cell. Biol. 8:3104-3113(1988).

CC -!- SUBCELLULAR LOCATION: NUCLEAR (Potential).
 CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M20678; AAA35990_1; -
 CC DR EMBL; M20679; ; NOT_ANNOTATED_CDS.
 DR PIR; F31201; F31201.
 DR HSP; P08046; 1A1C.
 DR Genew; HGNC;4320; GLI4.
 DR MIM; 105280; -.
 DR TrinPro; IPR000822; znf_C2H2.
 DR PRINTS; PR00048; ZINC_FINGER.
 DR SMART; SM00355; ZnF_C2H2; 6.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 6.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 6.
 KW ZINC-finger; Metal-binding; DNA-binding; Nuclear protein; Repeat.
 FT NON-TER 1 1
 DOMAIN 40 >129 ZINC FINGERS.
 FT ZN_FING 40 62 C2H2-TYPE.
 FT ZN_FING 68 90 C2H2-TYPE.
 FT ZN_FING 96 118 C2H2-TYPE.
 FT ZN_FING 124 >129 C2H2-TYPE.
 FT NON_CONS 129 130
 DOMAIN 220 <130 ZINC FINGERS.
 FT ZN_FING 136 136 C2H2-TYPE.
 FT ZN_FING 142 164 C2H2-TYPE.
 FT ZN_FING 170 192 C2H2-TYPE.
 FT ZN_FING 198 220 C2H2-TYPE.
 SQ SEQUENCE 223 AA; 25276 MW; 5DF24B508A04EA69 CRC64;
 Query Match 59.4%; Score 38; DB 1; Length 223;
 Best Local Similarity 63.6%; Pred. No. 6.7%;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 Qy | 1 TSPLNTHNGQK 11
 Db 112 ROHRINGER 122

RESULT 6

2.77_HUMAN ID FBX3_HUMAN
 ID 2177_HUMAN STANDARD; PRT; 321 AA.
 AC Q13360; 01-NOV-1997 (Rel. 3.5, Created)
 DT 01-NOV-1997 (Rel. 3.5, Last sequence update)
 DT 15-JUN-2002 (Rel. 4.1, Last annotation update)
 DE Zinc finger protein 177.

GN ZNPF177.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBITaxonID=9606;
 RN [1] SEQUENCE FROM N.A.
 RP MBDLINE=96229641; PubMed=8661005;
 RA Baban S., Freeman J.D., Mager D.L.;
 RT "Transcripts from a novel human KRAB zinc finger gene contain spliced
 Ali and endogenous retroviral segments.";
 RL Genomics 33:463-472 (1996).
 CC -!- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (Probable).

CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.

CC -!- SIMILARITY: CONTAINS 1 KRA3 DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; U37263; AAB09749_1; -
 DR HSSP; P08047; 1SF2.
 DR Genew; HGNC;2966; ZNF177.
 DR MIM; 601276; -.
 DR InterPro; IPR001909; KRB.
 DR DR PROSITE; IPR000822; Znf_C2H2.
 DR DR Pfam; PF00096; zf-C2H2; 7.
 DR DR Pfam; PF01352; KRAB; 1.
 DR DR PRINTS; PR00048; ZINC_FINGER.
 DR DR PRODOM; PD000003; Znf_C2H2; 7.
 DR DR SMART; SM00349; KRAB; 1.
 DR DR SMART; SM00355; Znf_C2H2; 7.
 DR DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
 DR DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 7.
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 DOMAIN 14 84 KRB.
 FT DOMAIN 124 314 ZINC FINGERS.
 FT ZN_FING 124 146 C2H2-TYPE.
 FT ZN_FING 152 174 C2H2-TYPE.
 FT ZN_FING 180 202 C2H2-TYPE.
 FT ZN_FING 208 230 C2H2-TYPE.
 FT ZN_FING 236 258 C2H2-TYPE.
 FT ZN_FING 264 286 C2H2-TYPE.
 FT ZN_FING 292 314 C2H2-TYPE.
 SQ SEQUENCE 321 AA; 36473 MW; E3258606C292DA17 CRC64;
 Query Match 59.4%; Score 38; DB 1; Length 321;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy | 6 IHNGQKL 12
 Db 313 IHNGQKL 319

RESULT 7

FBX3_HUMAN ID FBX3_HUMAN STANDARD; PRT; 471 AA.
 AC Q9UK99; Q9NUX2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE F-box only protein 3.
 FBX3 OR FBX3.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBITaxonID=9606;
 RN [1] SEQUENCE FROM N.A.
 RC TISSUE=Placenta.
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Matsunawa T., Nagai K., Sugano S., Aotsuka S., Yoshioka Y.,
 RA Nakamura Y., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
 RA "NEDO Human cDNA sequencing project.";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
 RN [2] SEQUENCE OF 6-415 FROM N.A.

Qy	5 NIHNGOKL 12	DT	01-OCT-1996 (Rel. 34, Last sequence update)
Db	117 NIHNGAKI 124	DT	15-JUN-2002 (Rel. 41, Last annotation update)
Qy	5 NIHNGOKL 12	DE	Hypothetical protein yggJ.
Db	117 NIHNGAKI 124	GN	YGGJ OR B2946.
Qy	5 NIHNGOKL 12	OS	Escherichia coli.
Db	117 NIHNGAKI 124	OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Qy	5 NIHNGOKL 12	OC	Escherichia.
Db	117 NIHNGAKI 124	OX	NEBIL_TaxID=562;
Qy	5 NIHNGOKL 12	RN	[1]
Db	117 NIHNGAKI 124	RN	SEQUENCE FROM N_A.
Qy	5 NIHNGOKL 12	RP	STRAIN=K12 / MG1655;
Db	117 NIHNGAKI 124	RP	BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V., RILEY M., COLLAZO-VIDES J., GLASNER J.D., RODE C.K., MAHONEY G.F., GREGOR J., COILLES N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J., MAU B., SHAO Y.;
Qy	5 NIHNGOKL 12	RA	"The complete genome sequence of Escherichia coli K-12.";
Db	117 NIHNGAKI 124	RA	Science 277:1453-1474 (1997).
Qy	5 NIHNGOKL 12	RN	[2]
Db	117 NIHNGAKI 124	RN	SEQUENCE OF 1-128 FROM N_A.
Qy	5 NIHNGOKL 12	RP	STRAIN=K12;
Db	117 NIHNGAKI 124	RP	JEKEL M., WACKERNAGEL W.;
Qy	5 NIHNGOKL 12	RA	"Location of the ends gene coding for endonuclease I on the physical map of the Escherichia coli K-12 chromosome.";
Db	117 NIHNGAKI 124	RA	J. Bacteriol. 176:1550-1551 (1994).
Qy	5 NIHNGOKL 12	RN	[3]
Db	117 NIHNGAKI 124	RN	SEQUENCE OF 127-243 FROM N_A.
Qy	5 NIHNGOKL 12	RP	STRAIN=B;
Db	117 NIHNGAKI 124	RP	MEDLINE=85087938; PubMed=6393055;
Qy	5 NIHNGOKL 12	RA	GUSHIMA H., YASUDA S., SODEA E., YOKOTA M., KONDO M., KIMURA A.;
Db	117 NIHNGAKI 124	RA	"Complete nucleotide sequence of the E. coli glutathione synthetase gene.";
Qy	5 NIHNGOKL 12	RT	Nucleic Acids Res. 12:9299-9307 (1984).
Qy	5 NIHNGOKL 12	RN	[4]
Db	117 NIHNGAKI 124	RP	IDENTIFICATION.
Qy	5 NIHNGOKL 12	RA	RIDD K.E., BAUM B.;
Db	117 NIHNGAKI 124	RL	Unpublished observations (AUG-1994);
Qy	5 NIHNGOKL 12	CC	-1- SIMILARITY: BELONGS TO THE UPF008 FAMILY. STRONG, TO H. INFLUENZAE H10303.
Db	117 NIHNGAKI 124	CC	-----
Qy	5 NIHNGOKL 12	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
Db	117 NIHNGAKI 124	CC	-----
Qy	5 NIHNGOKL 12	DR	EMBL; U28377; AA69113_1; ALT_INIT.
Db	117 NIHNGAKI 124	DR	EMBL; AE00377; AAC75983_1; ALT_INIT.
Qy	5 NIHNGOKL 12	DR	EMBL; X65169; ; NOT_ANNOTATED_CDS.
Db	117 NIHNGAKI 124	DR	EMBL; X01666; ; NOT_ANNOTATED_CDS.
Qy	5 NIHNGOKL 12	DR	EcoGene; EG1236; Y99J.
Db	117 NIHNGAKI 124	DR	InterPro; IPR003382; Cons_hypoth46.
Qy	5 NIHNGOKL 12	DR	Proteins; TIGR0046; Cons_hypoth46; 1.
Db	117 NIHNGAKI 124	DR	Hypothetical protein; Complete Proteome.
Qy	5 NIHNGOKL 12	SQ	SEQUENCE 243 AA; 26978 MW; 5C4659F5B295E033 CRC64;
Qy	5 NIHNGOKL 12	Qy	Query Match 56.2%; Score 36; DB 1; Length 195; Best Local Similarity 75.0%; Pred. No. 14; Mismatches 6; Conservative 1; Indels 0; Gaps 0;
Db	117 NIHNGAKI 124	Qy	Query Match 56.2%; Score 36; DB 1; Length 243; Best Local Similarity 77.8%; Pred. No. 17; Mismatches 7; Conservative 1; Indels 0; Gaps 0;
Qy	5 NIHNGOKL 12	Db	2 SPINTHNGQ 10
Qy	5 NIHNGOKL 12	Db	76 SPVHHLGQ 84
RESULT 15			Search completed: January 3, 2003, 14:25:08
YGGJ_ECOLI			Job time : 12 secs
ID YGGJ_ECOLI			
AC P37912; P76647; STANDARD;			
DT 01:OCT-1994 (Rel. 30, Created)			

Best Local Similarity 63.6%; Pred. No. 13; Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TSPLNTHNGOK 11
Oy | ||||: Db 112 TQHLRTHNGEK 122

RESULT 13

T24465 hypothetical protein T04FB.1 - *Caenorhabditis elegans*
C;date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T24465
R;Leonard, N.
submitted to the EMBL Data Library, November 1995
A;Reference number: 219895
A;Accession: T24465
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-641 <LET>
A;Cross-references: EMBL:U51999; PIDN:CAA91477.1; GSPDB:GN00028; CESP:T04FB.1
A;Experimental source: clone T04FB
C;Genetics:
A;Gene: CESP:T04FB.1
A;Map position: X
A;Introns: 29/3; 58/2; 172/3; 201/2; 244/1
C;Superfamily: Saccharomyces probable membrane protein YOR271C

Query Match 59.4%; Score 38; DB 2; Length 324; Best Local Similarity 54.5%; Pred. No. 21; Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TSPLNTHNGOK 11
Oy | :|||: :||| Db 35 TNPLNIFHGER 45

RESULT 14

AD0833 probable membrane protein STY2859 [imported] - *Salmonella enterica* subsp. *enterica* serovar Typhi C;Species: *Salmonella enterica* subsp. *enterica* serovar Typhi
A;Note: this species has also been called *Salmonella typhi*
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C;Accession: AD0833
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, S.; Threlfall, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moulé, S.; O'Gaora, P.
Nature 413, 848-852, 2001.
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi C;Reference number: AB0502; PMID:11677608
A;Accession: AD0833
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-406 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD05851.1; PID:g16503826; GSPDB:GN00176
C;Genetics:
A;Gene: STY2859

Query Match 59.4%; Score 38; DB 2; Length 406; Best Local Similarity 50.0%; Pred. No. 27; Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 3 PLNINHQQL 12
Oy | :|||: Db 356 PPRELHNGQRT 365

RESULT 15

T29991 hypothetical protein C43H6.4 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T29991
R;Le, T.T.
submitted to the EMBL Data Library, March 1996
A;Reference number: 220717
A;Accession: T29991
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Gene: CESP:C43H6.4
A;Map Position: X
A;Introns: 12/1; 63/2; 89/3; 143/1; 187/3; 268/3; 364/3; 403/3; 442/2; 487/2; 517/1
C;Superfamily: *Caenorhabditis elegans* hypothetical protein C43H6.4
Query Match 59.4%; Score 38; DB 2; Length 611; Best Local Similarity 72.7%; Pred. No. 45; Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 2 SPININHQQL 12
Db 510 SPININFKGYL 520

Search completed: January 3, 2003, 14:25:30
Job time : 16 secs

Best Local Similarity 58.3%; Pred. No. 3.2; Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0; Gaps 0; Gaps 0;	
QY	1 TSPLNLIHGQKL 12 1:11:1: 11
Db	122 TAPLNIVGGTKL 133
RESULT 3	
A64905	YdeK protein - Escherichia coli (strain K-12)
C;Species	Escherichia coli
C;Alternate names	protein T
C;Date	12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C;Accession	A64905; 152440; S4315
R;Blitner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co	A.; Rose, D.J.; Mau, B.; Shao, Y.
A;Title	The complete genome sequence of Escherichia coli K-12.
A;Reference number	A04720; MUID:97426617; PMID:9278503
A;Accession	A64905
A;Status	nucleic acid sequence not shown; translation not shown
A;Molecule type	DNA
A;Residues	1-1325 <BLAT>
A;Cross-references	GB:AE000248; GB:U00095; NID:gi1787783; PIDN:AC74583.1; PID:gi1787788;
A;Experimental source	strain K-12, substrain MG1655
R;Cartwright, P.; Timms, M.; Lithgow, T.; Hoj, P.; Hoogenraad, N.	Biochim. Biophys. Acta 1153, 345-347, 1993
A;Title	An Escherichia coli gene showing a potential ancestral relationship to the gene
A;Reference number	152440
A;Accession	152440
A;Status	translated from GB/EMBL/DDBJ
A;Molecule type	DNA
A;Residues	689-883, 'K', 885-1316, 'S', 1318-1325 <RES>
A;Cross-references	EMBL:X73295; NID:gi31292; PIDN:CAAS1730.1; PID:gi31293
A;Note	the difference in length is due to a frameshift error at pos 653
C;Genetics:	
A;Gee: ydeK	
C;Function:	
A;Description	probably involved in protein translocation apparatus
C;Keywords	nucleotide binding; P-loop
Query Match	Best Local Similarity 65.6%; Score 42; DB 2; Length 1325; Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	4 LNIHNGOKL 12 1 1 1 1 1 1
Db	991 LNIHGGOKL 999
RESULT 4	
E00893	hypothetical protein ECs2117 [imported] - Escherichia coli (strain 0157:H7, substrain R1)
C;Species	Escherichia coli
C;Date	18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession	E00893
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.	DNA Res. 8, 11-22, 2001
A;Title	Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7 and geno
A;Reference number	A99629; MUID:21156231; PMID:11258796
A;Accession	E00893
A;Status	preliminary
A;Molecule type	DNA
A;Residues	1-1343 <HAY>
A;Cross-references	GB:BA000007; PIDN:BAR35540.1; PID:gi13361583; GSPDB:GN00154
A;Experimental source	strain 0157:H7, substrain R1MD 0509952
C;Genetics:	
A;Gene	ECs2117
Query Match	65.6%; Score 42; DB 2; Length 1343;
QY	65.6%; Score 42; DB 2; Length 1343;
Db	65.6%; Score 42; DB 2; Length 1343;

Best Local Similarity 88.9%; Pred. No. 18; Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0; Gaps 0; Gaps 0;	
QY	4 LNIHNGOKL 12 1 1 1 1 1 1
Db	991 LNIHGGOKL 999
RESULT 5	
D85724	hypothetical protein ydeK [imported] - Escherichia coli (strain 0157:H7, substrain ED)
C;Species	Escherichia coli
C;Date	16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession	D85724
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayiller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Polamaisis, K.; Apoda	Nature 409, 529-533, 2001
A;Title	Genome sequence of enterohemorrhagic Escherichia coli 0157:H7.
A;Reference number	A65480; MUID:21074935; PMID:1126551
A;Accession	D85724
A;Status	preliminary
A;Molecule type	DNA
A;Residues	1-1343 <STO>
A;Cross-references	GB:AE005174; NID:gi12515159; PIDN:AAG56256.1; GSPDB:GN00145; WGDP: A;Experimental source: strain 0157:H7, substrain EDL933
C;Genetics:	
A;Gene	ydeK
Query Match	65.6%; Score 42; DB 2; Length 1343; Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	4 LNIHNGOKL 12 1 1 1 1 1 1
Db	991 LNIHGGOKL 999
RESULT 6	
T31070	notch homolog - sea urchin (Lytechinus variegatus)
C;Species	Lytechinus variegatus (variegated urchin)
C;Accession	T31070
R;Sherwood, D.R.; McClay, D.R.; Development, 124, 3363-3374, 1997	Development 124, 3363-3374, 1997
A;Title	Identification and localization of a sea urchin Notch homologue: insights in
A;Reference number	z20966; MUID:97454256; PMID:9310331
A;Accession	T31070
A;Status	preliminary; translated from GB/EMBL/DDBJ
A;Molecule type	mRNA
A;Residues	1-2531 <SHE>
A;Cross-references	EMBL:AF000634; NID:gi2570350; PIDN:92570351; PIDN:ABR82088.1
C;Superfamily	notch protein; ankyrin repeat homology; EGF homology
Query Match	65.6%; Score 42; DB 2; Length 2531; Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY	1 TSPLNLIHGQ 10 1 1 1 1 1 1
Db	2097 TSPMDMHNGE 2106
RESULT 7	
T08980	hypothetical protein f6G3.70 - Arabidopsis thaliana
C;Species	Arabidopsis thaliana (mouse-ear cress)
C;Date	11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999
C;Accession	T08980
R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.	submitted to the Protein Sequence Database, May 1999
A;Reference number	Z16520
A;Accession	T08980
Query Match	65.6%; Score 42; DB 2; Length 1343;

Run on:	January 3, 2003, 14:23:37	Search time 15 Seconds (without alignments)
Scoring table:	BLOSUM62	76.908 Million cell updates/sec
Searcher:	Gapext 10.0 , Gapext 0.5	
Total number of hits satisfying chosen parameters:	283224	
Minimum DB seq length:	0	
Maximum DB seq length:	2000000000	
Post-processing:	Minimum Match 0%	
	Maximum Match 100%	
Database :	PIR;73;*	
	1: pir1;*	
	2: pir2;*	
	3: pir3;*	
	4: pir4;*	
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.		
SUMMARIES		
Result No.	Score	Query Match Length DB ID
1	44	68.8 1302 2 B41249
2	42	65.6 300 2 S41535
3	42	65.6 1325 2 A64093
4	42	65.6 1343 2 E90893
5	42	65.6 1343 2 D85724
6	42	65.6 2531 2 T31070
7	41	64.1 427 2 T08980
8	40	62.5 331 2 T38246
9	39	60.9 243 2 AF0114
10	39	60.9 563 2 T36580
11	38	59.4 150 2 F86635
12	38	59.4 223 2 F13201
13	38	59.4 324 2 T24465
14	38	59.4 406 2 AD0833
15	38	59.4 641 2 T29951
16	38	59.4 683 2 T38254
17	38	57.8 693 2 C84495
18	37	57.8 111 2 111 2 G91129
19	37	57.8 111 2 G85974
20	37	57.8 111 2 111 2 111 2
21	37	57.8 240 2 F83592
22	37	57.8 248 2 S23348
23	37	57.8 299 2 S15300
24	37	57.8 57.8 111 2 AE0767
25	37	57.8 476 1 FOLBT
26	37	57.8 587 2 S58319
27	37	57.8 838 2 T04935
28	36	56.2 169 2 G83075
29	36	H71939
ALIGNMENTS		
RESULT 1		
B41249		
C;Species: Drosophila melanogaster		
C;Date: 03-Apr-1992 #sequence_revision 12-Jun-1992 #text_change 02-Feb-2001		
C;Accession: B41249		
R;Wu, C.T.; Budding, M.; Griffin, M.S.; Croop, J.M.		
Mol. Cell. Biol. 11, 3940-3948, 1991		
A;Title: Isolation and characterization of Drosophila multidrug resistance gene homolog		
A;Reference number: A41249; PMID:91304385; PMID:2072901		
A;Accession: B41249		
A;Status: Preliminary		
A;Molecule type: mRNA		
A;Residues: 1-1302 <WUA>		
A;Cross-references: GB:MS9077; NID:gi157874; PIDN:AAA28680_1; PID:gi157875		
C;Genetics:		
A;Gene: FlyBase:Md65		
A;Cross-references: FlyBase:FBgn0004513		
C;Superfamily: unassigned ATP-binding cassette proteins: ATP-binding cassette homolog		
C;Keywords: AMP; nucleotide binding; P-loop; transmembrane protein		
F;423-617//Domain: AMP-binding cassette homology <ABC1>		
F;1077-1274//Domain: ATP-binding cassette homology <ABC2>		
F;1094-1101//Region: nucleotide-binding motif A (P-loop)		
Query Match 68.8%; Score 44; DB 2; Length 1302;		
Best Local Similarity 72.7%; Pred. NO. 7.1;		
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;		
Qy 1 TSPINIHNGQK 11		
Db 673 TSPINLEKQK 683		
RESULT 2		
S41535		
C;Species: Shigella flexneri		
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 22-Jun-1999		
C;Accession: S41535		
R;Macpherson, D.F.; Manning, P.A.; Morona, R.		
Mol. Microbiol. 11, 281-292, 1994		
A;Title: Characterization of the dtDP-rihamnose biosynthetic genes encoded in the rfb		
A;Reference number: S41533; PMID:94224146; PMID:8170590		
A;Accession: S41535		
A;Status: preliminary		
A;Molecule type: DNA		
A;Residues: 1-300 <MAC>		
A;Cross-references: EMBL:X71970; NID:gi506557; PIDN:CAA50768_1; PID:gi454899		
C;Superfamily: dtDP-dihydrostreptose synthase		
Query Match 65.6%; Score 42; DB 2; Length 300;		

copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: January 3, 2003, 14:22:12 ; Search time 35 Seconds
(without alignments)

Scoring table: BIOSUM62

Scoring table: Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDS2/gcgdata/geneseq/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	%	Match Length	DB ID	Description
1	64	100.0	12	23	AA50916	Tumour infiltratin
2	64	100.0	19	23	AA50917	Tumour infiltratin
3	64	100.0	19	23	AA50918	Tumour infiltratin
4	42	65.6	804	22	ABG28800	Novel human diagno
5	42	65.6	804	22	ABG29023	Novel human diagno
6	42	65.6	807	22	ABG28665	Novel human diagno
7	42	65.6	1023	22	ABG26668	Novel human diagno
8	42	65.6	1302	22	ABG64919	Drosophila melanog
9	42	65.6	1325	22	ABG98222	Escherichia coli p
10	40	62.5	61	22	ABR41576	Peptide #9082 enco

SUMMARIES

No. 15 is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Escherichia coli β -peptide #9082 eno

PS Claim 2; Page 71; 104pp; English.

XX
XX
CC The present sequence is that of a synthetic peptide, termed HN-1,
CC that is specifically internalised by human head and neck squamous
CC carcinoma cells (HNSCC), or certain other solid tumour tissue cells,
CC such as breast cancer cells. HN-1 was identified by screening a
CC phage M13 peptide library displaying over 10 power 9 peptides. The
CC screening method was based on the ability of HNSCC line MDA167Tu
CC cells to uptake peptides by endocytosis at 37 degrees C. The cells
CC exhibited an approximately 10-fold greater internalisation
CC potential for HN-1 than normal human fibroblasts. The peptide
CC localised in cytoplasm after entry. *In vivo*, *i.v.* injected HN-1
CC peptide localised to HNSCC xenograft formed in nude mice. Claimed
CC compositions comprise HN-1 and a chemotherapeutic, cytotoxic,
CC apoptotic or DNA-damaging drug, such as taxol. The compositions
CC are used in claimed methods for killing a tumour cell, especially
CC an HNSCC (oral cavity, pharynx, throat, paranasal sinus, nasal
CC cavity, larynx, thyroid, parathyroid, salivary gland, face or neck
CC skin or cervical lymph node cell), breast cancer cell or other
CC solid tumour cell. Also claimed are: methods for detecting cancer
CC imaging and tumour treating kits; methods for killing tumour cells
CC in which a composition comprising an antitumour compound conjugated
CC to HN-1 is administered with radiotherapy, chemotherapy, surgery or
CC a gene therapy composition; and a method for the isolating of an
CC internalising peptide by phage display library screening. The
CC peptide provides the necessary dose of a drug specifically to the
CC tumours, avoiding harmful side effects on other cells. The peptide
CC is non-toxic, non immunogenic, stable *in vivo*, protects its cargo
CC during transit, and accumulates in a tumour within 48 hours.
XX
SQ Sequence 12 AA;

Query Match 100.0%; Score 64; DB 23; Length 12;
Best Local Similarity 100.0%; Pred. No. 1e-05; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPLNIGHQQL 12
||| | | | | | | |
Db 1 TSPLNIGHQQL 12

RESULT²
ID AAM50917

ID AAM50917 standard; Peptide; 19 AA.

AC AAM50917;

XX DT 07-MAY-2002 (first entry)

XX DE Tumour infiltrating peptide HN-2.

XX KW Tumour infiltrating peptide; HN-2; head and neck cancer; HNSCC;
KW breast cancer; drug delivery; diagnosis; imaging; gene therapy.
XX OS Synthetic.

XX FR Key Location/Qualifiers

FT Peptide 4.15
/note= "corresponds to HN-1"

PN WO200202147-A2.

XX PD 10-JAN-2002.

XX PF 02-JUL-2001; 2001WO-US21518.

XX PR 30-JUN-2000; 2000US-215491P.

XX PA (TEXA) UNIV TEXAS SYSTEM.

XX PI Clayman G, Hong FD;

XX DR WPI; 2002-195737/25.

XX
XX
PT Peptide internalised by a tumour cell useful for targeted delivery of
PT anticancer drugs -
XX
PS Example 2; Page 76; 104pp; English.
XX
CC The present sequence is that of a synthetic peptide, termed HN-2,
CC which is based on the claimed tumour infiltrating peptide HN-1 (see
CC AAM50916) with additional N- and C-terminal amino acid residues.
CC HN-1 is specifically internalised by human head and neck squamous
CC carcinoma cells (HNSCC) and certain other solid tumour tissue cells,
CC such as breast cancer cells. The additional amino acid residues of
CC HN-2 did not inhibit cell internalisation of the peptide; HN-1
CC internalisation is position-independent but sequence-dependent.
CC Claimed compositions comprise HN-1 and a chemotherapeutic, cytotoxic,
CC apoptotic or DNA-damaging drug, such as taxol. The compositions
CC are used in claimed methods for: killing a tumour cell, especially
CC an HNSCC (oral cavity, pharynx, throat, paranasal sinus, nasal
CC cavity, larynx, thyroid, parathyroid, salivary gland, face or neck
CC skin or cervical lymph node cell), breast cancer cell or other
CC solid tumour cell. Also claimed are: methods for detecting cancer
CC imaging and tumour treating kits; methods for killing tumour cells
CC in which a composition comprising an antitumour compound conjugated
CC to HN-1 is administered with radiotherapy, chemotherapy, surgery or
CC a gene therapy composition; and a method for the isolating of an
CC internalising peptide by phage display library screening. The
CC peptide provides the necessary dose of a drug specifically to the
CC tumours, avoiding harmful side effects on other cells. The peptide
CC is non-toxic, non immunogenic, stable *in vivo*, protects its cargo
CC during transit, and accumulates in a tumour within 48 hours.
XX
SQ Sequence 19 AA;

Query Match 100.0%; Score 64; DB 23; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.8e-05; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPLNIGHQQL 12
||| | | | | | | |
Db 4 TSPLNIGHQQL 15

RESULT³
ID AAM50918

ID AAM50918 standard; Peptide; 19 AA.

AC AAM50918;

XX DT 07-MAY-2002 (first entry)

XX DE Tumour infiltrating peptide HN-3.

XX KW Tumour infiltrating peptide; HN-3; head and neck cancer; HNSCC;
KW breast cancer; drug delivery; diagnosis; imaging; gene therapy.
XX OS Synthetic.

XX FR Key Location/Qualifiers

FT Peptide 8.19
/note= "corresponds to HN-1"

PN WO200202147-A2.

XX PD 10-JAN-2002.

XX PF 02-JUL-2001; 2001WO-US21518.

XX PR 30-JUN-2000; 2000US-215491P.

XX PA (TEXA) UNIV TEXAS SYSTEM.

XX PI Clayman G, Hong FD;

XX DR WPI; 2002-195737/25.

XX
DR
XX
WPI: 2002-195737/25.

XX
PT peptide internalised by a tumour cell useful for targeted delivery of
XX
anticancer drugs -

PS Example 2; Page 76; 104pp; English.

XX
CC The present sequence is that of a synthetic peptide, termed HN-1 (see
CC which is based on the claimed tumour infiltrating peptide HN-1 (see
XX
ARM0316) with additional N-terminal amino acid residues. HN-1 is
CC specifically internalised by human head and neck squamous carcinoma
CC cells (HNSCC) and certain other solid tumour tissue cells, such as
CC breast cancer cells. The additional N-terminal amino acid residues
CC of HN-3 did not inhibit cell internalisation of the peptide; HN-1
CC internalisation is position-independent but sequence-dependent.
CC
CC claimed compositions comprise HN-1 and a chemotherapeutic, cytotoxic,
CC apoptotic or DNA-damaging drug, such as taxol. The compositions
CC are used in claimed methods for killing a tumour cell, especially
CC an HNSCC (oral cavity, pharynx, throat, paranasal sinus, nasal
CC cavity, larynx, thyroid, parathyroid, salivary gland, face or neck
CC skin or cervical lymph node cell), breast cancer cell or other
CC solid tumour cell. Also claimed are: methods for detecting cancer
CC by labeling HN-1 with a detectable label; tumour detection, tumour
CC imaging and tumour treating kits; methods for killing tumour cells
CC in which a composition comprising an antitumour compound conjugated
CC to HN-1 is administered with radiotherapy, chemotherapy, surgery or
CC a gene therapy composition; and a method for the isolating of an
CC internalising peptide by phage display library screening. The
CC peptide provides the necessary dose of a drug specifically to the
CC tumours, avoiding harmful side effects on other cells. The peptide
CC is non-toxic, non-immunogenic, stable in vivo, protects its cargo
CC during transit, and accumulates in a tumour within 48 hours.

XX
SQ Sequence 19 AA;

Query Match 100.0%; Score 64; DB 23; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TSPUNTHNGQKL 12
Db 8 TSPLNTHNGQKL 19

RESULT 4
ABG28800 standard; Protein: 804 AA.

XX
AC ABG28800;
XX
DT 18-FEB-2002 (first entry)

XX
DE Novel human diagnostic protein #28791.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.

XX
PN WO200175067-A2.

XX
PD 11-OCT-2001.

XX
PR 30-MAR-2001; 2001WO-US08631.

XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.

XX
PA (HYSE-) HYSEQ INC.

XX
PI Drmanac RT, Liu C, Tang YT;

XX
DR N-PSDB; AAS93210.

DR
WPI: 2001-639362/73.
N-PSDB; AAS92987.

XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID No 59159; 103pp; English.

XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II); (II) is useful for generating antibodies against it; detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp://wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 804 AA;

Query Match 65.6%; Score 42; DB 22; Length 804;
Best Local Similarity 88.9%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 4 LNTHNGQKL 12
Db 325 LNTHGGQKL 333

RESULT 5

ABG29023
ID ABG29023 standard; Protein: 804 AA.

XX
AC ABG29023;
XX
DT 18-FEB-2002 (first entry)

XX
DE Novel human diagnostic protein #29014.

XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.

XX
PD 11-OCT-2001.

XX
PR 30-MAR-2001; 2001WO-US08631.

XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.

XX
PA (HYSE-) HYSEQ INC.

XX
PI Drmanac RT, Liu C, Tang YT;

XX
DR N-PSDB; AAS93210.

XX	PT	New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -
XX	CC	Claim 20; SEQ ID No 59382; 103pp; English.
XX	CC	The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy technique to restore normal activity of (II) or to treat disease states involving (I). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (III) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences.
XX	CC	ABG0010-ABG3037 represent novel human diagnostic amino acid sequences of the invention.
XX	CC	Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at http://wipo.int/pub/published_pct_sequences .
XX	SQ	Sequence 804 AA:
XX		Query Match 65.6%; Score 42; DB 22; Length 804; Best Local Similarity 88.9%; Pred. NO. 25; Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0
QY		4 LNIHNGOKL 12
		325 LNIHGGOKL 333
RESULT	6	
ID	ABG28665	Arg28665 standard; Protein: 807 AA.
XX	ABG28665:	
AC		
XX		
DT	18-FEB-2002	(first entry)
XX		
DE	Novel human diagnostic protein #28656.	
XX		
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200175067-A2.	
XX		
PD	11-OCT-2001.	
XX		
PF	30-MAR-2001; 2001WO-US08631.	
XX		
PR	31-MAR-2000; 2000US-0540217.	
XX		
PR	23-AUG-2000; 2000US-0649167.	
XX		
PA	(HYSE-) HYSEQ INC.	
XX		
PT	Drimanac RT, Liu C, Tang YT;	
XX		
DR	WPI; 2001-639362/73.	
XX		
N-PSDB; AAS92852.		
XX		
PT	New isolated polynucleotide and encoded polypeptides, useful in	

PT biodiversity -
 XX
 PS Claim 20; SEQ ID NO 59027; 103pp; English.
 XX
 The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC and gene mapping, and in recombinant production of (II). The
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. (ABG0010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp://wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1023 AA;
 Query Match 65.6%; Score 42; DB 22; Length 1023;
 Best Local Similarity 88.9%; Pred. No. 33; Mismatches 0;
 Matches 8; Conservative 0; Indels 1; Gaps 0;
 QY 4 LNIHNGQKL 12
 YY 11111111
 Db 439 LNIHGQKL 447
 AC ABB64919;
 XX
 RESULT 8
 ABB64919
 ID ABB64919 standard; Protein; 1302 AA.
 XX
 AC ABB64919;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 21549.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 XX
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO20011042-A2.
 XX
 PD 05-JUL-2001.
 XX
 PF 19-DEC-2000; 2000WO-US34419.
 XX
 PR 23-DEC-1999; 990US-0173005.
 XX
 PA (ELIT-) ELTRA PHARM INC.
 XX
 PT Forsyth RA, Ohlsen KL, Zyskind JW;
 XX
 DR WPI; 2001-457376/49.
 DR N-PSDB; AAH81312.
 XX
 PT Novel nucleic acids encoding proteins required for *Escherichia coli*
 XX proliferation, useful for screening for antimicrobial agents -
 XX
 PS Claim 19; Page 445-448; 596pp; English.
 XX
 CC the present invention describes a purified or isolated nucleic acid
 CC sequence (I) consisting essentially of one of the 93 nucleotide sequences
 CC given in AAH81202 to AAH81294, where expression of the nucleic acid in a
 CC microorganism is capable of inhibiting proliferation of a microorganism.
 CC (I) have antibacterial and antibiotic activities, and can be used in
 CC gene therapy. Expression of (I) in a microorganism inhibits proliferation
 CC of the microorganism, and the manufactured antibiotic is useful for
 CC reducing the activity or level of a gene product required for
 CC proliferation of a microorganism in a subject, specifically humans. The
 CC nucleic acids that inhibit bacterial growth or proliferation can be used
 CC as antisense therapeutics for killing bacteria. In addition to
 CC therapeutic applications, the nucleic acid sequences complementary to
 CC sequences required for proliferation can be used as diagnostic tools.

XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176 ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABL57737-ABL72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp://wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1302 AA;
 Query Match 65.6%; Score 42; DB 22; Length 1302;
 Best Local Similarity 72.7%; Pred. No. 44; Mismatches 0;
 Matches 8; Conservative 0; Indels 3; Gaps 0;
 QY 1 TSPLNIHNGQKL 11
 YY 11111111
 Db 673 TSPLNFEKGQK 683
 AC AAG98256
 XX
 RESULT 9
 AAG98256 standard; Protein; 1325 AA.
 XX
 AC AAG98256;
 XX
 DT 21-SEP-2001 (first entry)
 XX
 DE *Escherichia coli* protein sequence SEQ ID NO:304.
 XX
 KW *Escherichia coli*; identification; proliferation; microorganism;
 KW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
 KW bacterial growth inhibition.
 XX
 OS *Escherichia coli*.
 XX
 PN WO200148209-A2.
 XX
 PD 05-JUL-2001.
 XX
 PF 19-DEC-2000; 2000WO-US34419.
 XX
 PR 23-DEC-1999; 990US-0173005.
 XX
 PA (ELIT-) ELTRA PHARM INC.
 XX
 PT Forsyth RA, Ohlsen KL, Zyskind JW;
 XX
 DR WPI; 2001-457376/49.
 DR N-PSDB; AAH81312.
 XX
 PT Novel nucleic acids encoding proteins required for *Escherichia coli*
 XX proliferation, useful for screening for antimicrobial agents -
 XX
 PS Claim 19; Page 445-448; 596pp; English.
 XX
 CC the present invention describes a purified or isolated nucleic acid
 CC sequence (I) consisting essentially of one of the 93 nucleotide sequences
 CC given in AAH81202 to AAH81294, where expression of the nucleic acid in a
 CC microorganism is capable of inhibiting proliferation of a microorganism.
 CC (I) have antibacterial and antibiotic activities, and can be used in
 CC gene therapy. Expression of (I) in a microorganism inhibits proliferation
 CC of the microorganism, and the manufactured antibiotic is useful for
 CC reducing the activity or level of a gene product required for
 CC proliferation of a microorganism in a subject, specifically humans. The
 CC nucleic acids that inhibit bacterial growth or proliferation can be used
 CC as antisense therapeutics for killing bacteria. In addition to
 CC therapeutic applications, the nucleic acid sequences complementary to
 CC sequences required for proliferation can be used as diagnostic tools.

XX
 CC New isolated nucleic acid detection reagent for detecting 1000 or more
 CC genes from *Drosophila* and for elucidating cell signalling and cell-cell
 CC interactions -
 XX
 PS Disclosure; SEQ ID NO 21549; 21pp + Sequence Listing; English.

For example, nucleic acid probes complementary to proliferation-required sequences that are specific for particular species of microorganisms can be used as probes to identify particular microorganism species in clinical specimens. AAH81295 to AAH81487 encode the *Escherichia coli* proteins given in AAG8239 to AAG9831, and AAH81488 to AAH81491 represent oligonucleotides, which are used in the exemplification of the present invention.

RESULT 10
Query Match 65.6%; Score 42; DB 22; Length 1325;
Best Local Similarity 88.9%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 4 LNIHNGKL 12
Db 991 LNIHGGKL 999

RESULT 11
Query Match 65.6%; Score 42; DB 22; Length 1325;
Best Local Similarity 88.9%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 TSPLNHNGQK 11
Db 35 TSPKHKNGQK 45

RESULT 10
Query Match 65.6%; Score 42; DB 22; Length 1325;
Best Local Similarity 88.9%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 4 LNIHNGKL 12
Db 991 LNIHGGKL 999

RESULT 10
Query Match 65.6%; Score 42; DB 22; Length 1325;
Best Local Similarity 88.9%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 4 LNIHNGKL 12
Db 991 LNIHGGKL 999

04-FEB-2002 (first entry)
Peptide #9082 encoded by human foetal liver single exon probe.
KW Human; foetal liver; gene expression; single exon nucleic acid probe.
KW Homo sapiens.
XX WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00669.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 30-JUN-2000; 2000US-0207456.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0334687.
XX PR 27-SEP-2000; 2000US-0336359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-48889/53.
XX PS Claim 15; SEQ ID No 27186; 530pp; English.
XX CC The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart (see ABA21335-ABA41305). The present sequence is a protein encoded by one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosis diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.
XX CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 61 AA:
Query Match 62.5%; Score 40; DB 22; Length 61;
Best Local Similarity 72.7%; Pred. No. 3.1;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 TSPLNHNGQK 11
Db 35 TSPKHKNGQK 45

RESULT 12
Query Match 62.5%; Score 40; DB 22; Length 61;
Best Local Similarity 72.7%; Pred. No. 3.1;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 TSPLNHNGQK 11
Db 35 TSPKHKNGQK 45

XX	PS	Claim 27: SEQ ID NO 34569; 634pp: English.
PI	XX	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	CC	WPI: 2001-488897/53.
DR	CC	Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human placenta -
XX	CC	The present invention relates to single exon nucleic acid probes (SENPs: see AAI31315-AA157546). The present sequence is a peptide encoded by one such probe. The probes are useful for producing a microarray for predicting, measuring and displaying gene expression in samples derived from human placenta. The probes are useful for antenatal diagnosis of human genetic disorders.
PT	CC	Claim 27: SEQ ID NO 35637; 654pp: English.
XX	CC	The present invention relates to single exon nucleic acid probes (SENPs: see AAI31315-AA157546). The present sequence is a peptide encoded by one such probe. The probes are useful for producing a microarray for predicting, measuring and displaying gene expression in samples derived from human placenta. The probes are useful for antenatal diagnosis of human genetic disorders.
PS	CC	Sequence 61 AA;
XX	CC	Query Match 62.5%; Score 40; DB 22; Length 61;
CC	CC	Best local Similarity 72.7%; Pred. No. 3.1;
CC	CC	Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY	CC	1 TSPLNIGHNGQK 11
Db	CC	:
35	CC	TSPKHIKNGQK 45
RESULT 15	CC	ABG44904
ID	CC	ABG44904 standard; Peptide: 61 AA.
XX	CC	ABG44904;
AC	CC	19-AUG-2002 (first entry)
XX	CC	Human peptide encoded by genome-derived single exon probe SEQ ID 34569.
DE	CC	Human; single exon probe; asthma; lung cancer; COPD; ILD; familial obstructive pulmonary disease; interstitial lung disease; tuberous sclerosis; Gaucher's disease; Niemann-Pick disease; Hermansky-Pudlak syndrome; sarcoidosis; pulmonary histiocytosis; lymphangioleiomyomatosis; primary ciliary dyskinesia; pulmonary alveolar proteinosis; Karagener syndrome; pulmonary dysplasia; primary ciliary dyskinesia; pulmonary hypertension and hyaline membrane disease. The present sequence is a peptide/protein encoded by a single exon probe of the invention.
KW	CC	Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at fip.wipo.int/pub/published.pct.sequences.
KW	CC	Sequence 61 AA;
XX	CC	Query Match 62.5%; Score 40; DB 23; Length 61;
OS	CC	Best local Similarity 72.7%; Pred. No. 3.1;
OS	CC	Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
PN	CC	1 TSPLNIGHNGQK 11
XX	CC	:
PD	CC	35 TSPKHIKNGQK 45
15-NOV-2001.	CC	Search completed: January 3, 2003, 14:24:51
XX	CC	Job time : 37 secs
PP	CC	04-FEB-2000; 2000US-180312P.
PR	CC	26-MAY-2000; 2000US-207456P.
PR	CC	30-JUN-2000; 2000US-0008408.
PR	CC	03-AUG-2000; 2000US-0532365.
PR	CC	21-SEP-2000; 2000US-214687P.
PR	CC	27-SEP-2000; 2000US-236359P.
PR	CC	04-OCT-2000; 2000GB-0024263.
XX	CC	(MOLE-) MOLECULAR DYNAMICS INC.
PI	CC	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	CC	WPI: 2002-114183/15.
DR	CC	Spatially-addressable set of single exon nucleic acid probes, used to measure gene expression in human lung samples -
XX	CC	XX

OM protein - protein search, using sw model

Run on: January 3, 2003, 14:25:38 ; Search time 35 Seconds
(without alignments)
45.686 Million cell updates/sec

Title: US-09-899-376-1

Perfect score: 12

Sequence: 1 TSPINIHNGOKL 12

Scoringtable: Gapext 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_101002:*

1: /SIDS2/gcadata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDS2/gcadata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDS2/gcadata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDS2/gcadata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDS2/gcadata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDS2/gcadata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDS2/gcadata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDS2/gcadata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDS2/gcadata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDS2/gcadata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDS2/gcadata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDS2/gcadata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDS2/gcadata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDS2/gcadata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDS2/gcadata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDS2/gcadata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDS2/gcadata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDS2/gcadata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDS2/gcadata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDS2/gcadata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDS2/gcadata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDS2/gcadata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDS2/gcadata/geneseq/geneseq-emb1/AA2002.DAT:*

11 6 50.0 162 22 ABG24051 Novel human diagno
12 6 50.0 467 23 ABP8390 Staphylococcus epi
13 6 50.0 472 22 AAG2404 S. epidermidis ope
14 5 41.7 20 16 AAW13187 Fragment of p53 bi
15 5 41.7 20 19 AAW57358 Human WIP1 immunog
16 5 41.7 20 22 AAB6642 Human WIP1 tissue fact
17 5 41.7 39 21 AAB4615 Human secreted pro
18 5 41.7 45 20 AAW67843 Human secreted pro
19 5 41.7 45 22 AAB63358 Human breast cancer
20 5 41.7 52 19 AAW39901 Human plasma membran
21 5 41.7 54 22 AAM4431 Human immune/haema
22 5 41.7 59 19 AAW77369 Human WIP1 polyval
23 5 41.7 62 22 AAB0564 Propionibacterium
24 5 41.7 63 22 ABB67816 Drosophila melanog
25 5 41.7 68 22 AAU22368 Human cardiovascu
26 5 41.7 70 22 AAU5030 Human propionibacterium
27 5 41.7 70 22 AAQ12788 Human propionibacterium
28 5 41.7 72 21 AAG03674 Human polypeptide
29 5 41.7 72 22 ABB03553 Human secreted pro
30 5 41.7 80 22 AAQ08237 Human secreted pro
31 5 41.7 82 23 ABP04929 Human ORPX protein
32 5 41.7 86 22 AAM07048 Human ORPX protein
33 5 41.7 88 21 AAB43048 Human ORPX protein
34 5 41.7 89 22 AAU22146 Human cardiovascu
35 5 41.7 90 21 AAB08525 Human musculoskele
36 5 41.7 91 22 AAU2873 Human polypeptide
37 5 41.7 97 22 AAM03716 Novel human secret
38 5 41.7 99 18 AAW29891 Human EST encoded
39 5 41.7 99 18 AAW22074 BRCA-1 fragment 1'
40 5 41.7 100 20 AAY15267 Chlamydia pneumoniae
41 5 41.7 102 21 AAG51990 Arabidopsis thalia
42 5 41.7 103 21 AAG08049 Arabidopsis thalia
43 5 41.7 105 22 AAU42765 Propionibacterium
44 5 41.7 105 22 AAU20778 Human novel foetal
45 5 41.7 108 22 AAU09641 Human novel foetal
45 5 41.7 108 22 AAU09641 Human polypeptide

ALIGNMENTS

RESULT 1
ID AAM0916
AAM0916 standard; Peptide; 12 AA.
XX
AC AAM0916;
XX DT 07-MAY-2002 (first entry)
XX DE Tumour infiltrating peptide HN-1.
XX KW Tumour infiltrating peptide; HN-1; head and neck cancer; HNSCC;
KW breast cancer; drug delivery; diagnosis; imaging; gene therapy.
XX OS Synthetic.
XX PN WO200202147-A2.
XX PD 10-JAN-2002.
XX PR 02-JUL-2001; 2001WO-0S21518.
XX PR 30-JUN-2000; 2000US-215491P.

(TEXA) UNIV TEXAS SYSTEM.

PI Clayman G, Hong FD;

XX DR WPI; 2002-195737/25.

PT Peptide internalised by a tumour cell useful for targeted delivery of
anticancer drugs -

XX PS Claim 2; Page 71; 104pp; English.

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12	100.0	12	23 AAM50916	Tumour infiltratin
2	12	100.0	19	23 AAM50917	Tumour infiltratin
3	12	100.0	19	23 AAM50918	Tumour infiltratin
4	7	58.3	247	21 AAV83081	F-box protein FBP-
5	7	58.3	408	22 AAB48304	Human zr23 protein
6	7	58.3	471	21 AAV90287	Human peptidase, H
7	7	58.3	471	22 AAB53481	Human protein sequ
8	7	58.3	471	23 AAB47764	F-cassette structu
9	6	50.0	109	23 ABP28678	Staphylococcus epi
10	6	50.0	130	21 AAG37863	Arabidopsis thalia

XX
 CC The present sequence is that of a synthetic peptide, termed HN-1,
 CC that is specifically internalised by human head and neck squamous
 CC carcinoma cells (HNSCC), or certain other solid tumour tissue cells,
 CC such as breast cancer cells. HN-1 was identified by screening a
 CC phage M13 peptide library displaying over 10 power 9 peptides. The
 screening method was based on the ability of HNSCC line MDA167Tu
 CC cells to uptake peptides by endocytosis at 37 degrees C. The cells
 CC exhibited an approximately 10-fold greater internalisation. The peptide
 CC potential for HN-1 than normal human fibroblasts. The peptide
 CC localised in cytoplasm after entry. *In vivo*, i.v. injected HN-1
 peptide localised to HNSCC xenograft formed in nude mice. Claimed
 CC compositions comprise HN-1 and a chemotherapeutic, cytotoxic,
 CC apoptotic or DNA-damaging drug, such as taxol. The compositions
 CC are used in claimed methods for killing a tumour cell, especially
 CC an HNSCC (oral cavity, pharynx, throat, paranasal sinus, nasal
 CC cavity, larynx, thyroid, parathyroid, salivary gland, face or neck
 CC skin or cervical lymph node cell), breast cancer cell or other
 CC solid tumour cell. Also claimed are: methods for detecting cancer
 CC imaging and tumour treating kits; methods for killing tumour cells
 CC in which a composition comprising an antitumour compound conjugated
 CC to HN-1 is administered with radiotherapy, chemotherapy, surgery or
 CC a gene therapy composition; and a method for the isolating of an
 CC internalising peptide by phage display library screening. The
 CC peptide provides the necessary dose of a drug specifically to the
 CC tumours, avoiding harmful side effects on other cells. The peptide
 CC is non toxic, non-immunogenic, stable *in vivo*, protects its cargo
 XX during transit, and accumulates in a tumour within 48 hours.
 SQ Sequence 12 AA;

Query Match 100.0%; Score 12; DB 23; Length 12;
 Best Local Similarity 100.0%; Pred. No. 3.1e-07;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPLNIGHNGQKL 12
 1 ||||||| 12
 Db 1 TSPLNIGHNGQKL 12

RESULT 2
 ID AAM50917
 ID AAM50917 standard; Peptide; 19 AA.

AC AAM50917;
 AC
 DT 07-MAY-2002 (first entry)

DE Tumour infiltrating peptide HN-2.

XX
 KW Tumour infiltrating peptide; HN-2; head and neck cancer; HNSCC;
 KW breast cancer; drug delivery; diagnosis; imaging; gene therapy.
 XX
 OS Synthetic.

XX
 FH Key Location/Qualifiers
 FT Peptide 4..15
 /note= "corresponds to HN-1"

WO200202147-A2.

XX
 PD 10-JAN-2002.

XX
 PF 02-JUL-2001; 2001WO-US21518.

XX
 PR 30-JUN-2000; 2000US-215491P.

XX
 PA (TEXA) UNIV TEXAS SYSTEM.

XX
 PI Clayman G, Hong FD;

XX
 DR WO; 2002-195737/25.

XX
 PT Peptide internalised by a tumour cell useful for targeted delivery of
 PT anticancer drugs -

XX
 Example 2; Page 76; 104PP; English.

XX
 CC The present sequence is that of a synthetic peptide, termed HN-2,
 CC which is based on the claimed tumour infiltrating peptide HN-1 (see
 CC AAM50916) with additional N- and C-terminal amino acid residues.
 CC HN-1 is specifically internalised by human head and neck squamous
 CC carcinoma cells (HNSCC) and certain other solid tumour tissue cells,
 CC such as breast cancer cells. The additional amino acid residues of
 CC HN-2 did not inhibit cell internalisation of the peptide; HN-1
 CC internalisation is position-independent but sequence-dependent.
 CC Claimed compositions comprise HN-1 and a chemotherapeutic, cytotoxic,
 CC apoptotic or DNA-damaging drug, such as taxol. The compositions
 CC are used in claimed methods for killing a tumour cell, especially
 CC an HNSCC (oral cavity, pharynx, throat, paranasal sinus, nasal
 CC cavity, larynx, thyroid, parathyroid, salivary gland, face or neck
 CC skin or cervical lymph node cell), breast cancer cell or other
 CC solid tumour cell. Also claimed are: methods for detecting cancer
 CC imaging and tumour treating kits; methods for killing tumour cells
 CC in which a composition comprising an antitumour compound conjugated
 CC to HN-1 is administered with radiotherapy, chemotherapy, surgery or
 CC a gene therapy composition; and a method for the isolating of an
 CC internalising peptide by phage display library screening. The
 CC peptide provides the necessary dose of a drug specifically to the
 CC tumours, avoiding harmful side effects on other cells. The peptide
 CC is non-toxic, non-immunogenic, stable *in vivo*, protects its cargo
 CC during transit, and accumulates in a tumour within 48 hours.

XX
 SQ Sequence 19 AA;

Query Match 100.0%; Score 12; DB 23; Length 19;
 Best Local Similarity 100.0%; Pred. No. 4.8e-07;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSPLNIGHNGQKL 12
 1 ||||||| 19
 Db 4 TSPLNIGHNGQKL 15

RESULT 3
 ID AAM50918
 ID AAM50918 standard; Peptide; 19 AA.

AC AAM50918;
 AC
 DT 07-MAY-2002 (first entry)

DE Tumour infiltrating peptide HN-3.

XX
 KW Tumour infiltrating peptide; HN-3; head and neck cancer; HNSCC;
 KW breast cancer; drug delivery; diagnosis; imaging; gene therapy.
 XX
 OS Synthetic.

XX
 FH Key Location/Qualifiers
 FT Peptide 8..19
 /note= "corresponds to HN-1"

XX
 PN WO200202147-A2.

XX
 PD 10-JAN-2002.

XX
 PF 02-JUL-2001; 2001WO-US21518.

XX
 PR 30-JUN-2000; 2000US-215491P.

XX
 PA (TEXA) UNIV TEXAS SYSTEM.

XX
 PI Clayman G, Hong FD;

XX
XX
DR
WPI; 2002-195737/25.

XX
PT
Peptide internalised by a tumour cell useful for targeted delivery of
anticancer drugs -

PS
Example 2; Page 76; 104pp; English.

The present sequence is that of a synthetic peptide, termed HN-1 (see AAM5016) with additional N-terminal amino acid residues. HN-1 is specifically internalised by human head and neck squamous carcinoma cells (HNSCC) and certain other solid tumour tissue cells, such as breast cancer cells. The additional N-terminal amino acid residues of HN-3 did not inhibit cell internalisation of the peptide; HN-1 internalisation is position independent but sequence-dependent. Claimed compositions comprise HN-1 and a chemotherapeutic, cytotoxic, apoptotic or DNA-damaging drug, such as taxol. The compositions are used in claimed methods for killing a tumour cell, especially an HNSCC (oral cavity, pharynx, throat, paranasal sinus, nasal cavity, larynx, thyroid, parathyroid, salivary gland, face or neck skin or cervical lymph node cell), breast cancer cell or other solid tumour cell. Also claimed are: methods for detecting cancer by labeling HN-1 with a detectable label; tumour detection, tumour imaging and tumour treating kits; methods for killing tumour cells in which a composition comprising an antitumour compound conjugated to HN-1 is administered with radiotherapy, chemotherapy, surgery or a gene therapy composition; and method for the isolating of an peptide provides the necessary dose of a drug specifically to the tumours, avoiding harmful side effects on other cells. The peptide is non-toxic, non-immunogenic, stable in vivo, protects its cargo during transit, and accumulates in a tumour within 48 hours.

Sequence 19 AA:

Query Match 100.0%; Score 12; DB 23; Length 19;
Best Local Similarity 100.0%; Pred. No. 4.8e-07; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSPLNTHNGQKL 12

Db 8 TSPLNTHNGQKL 19

RESULT 4
AY83081
ID AY83081 standard; Protein: 247 AA.

AY83081;
XX
AC
XX
DT
XX
DE
XX
F-box protein FBP-13.

F-box protein; FBP; diagnosis; treatment; screening; agonist; antagonist; proliferative disorder; differentiative disorder; breast cancer; prostate cancer; ovarian cancer; cancer; small cell lung carcinoma; immune disorder; cardiovascular disorder; inflammatory disorder; human.

XX
OS
Homo sapiens.

XX
Homo sapiens.

F-box protein; FBP; diagnosis; treatment; screening; agonist; antagonist; proliferative disorder; differentiative disorder; breast cancer; prostate cancer; ovarian cancer; cancer; small cell lung carcinoma; immune disorder; cardiovascular disorder; inflammatory disorder; human.

XX
OS
Homo sapiens.

XX
WO200012679-A1.

PD
09-MAR-2000.

XX
PF
27-AUG-1999; 99WO-US19560.

XX
PR
28-AUG-1998; 98US-0098355.

PR
03-FEB-1999; 99US-0118568.

PR
15-MAR-1999; 99US-0124449.

PA (UNIV NEW YORK STATE.
XX
PI Chlauro DS, Pagano M, Latres E;
XX
DR WPI; 2000-256635/22.
XX
DR N-PSDB; AAC93363.

XX
PT
Novel nucleic acid for screening compounds useful for treating
proliferative and differentiative disorders such as cancer and immune
disorders comprises sequences encoding ubiquitin ligases -

PS
Claim 10; Figure 16a; 245pp; English.

XX
PT
Novel nucleic acid encoding substrate-targeting subunits of ubiquitin
ligases with F-box motifs (F-box proteins) are useful for diagnosis
of proliferative and differentiated related disorders by measuring
FBP gene expression. Cells expressing such proteins or
their fragments are useful for screening compounds. The compounds
are agonists or antagonists, which are useful for treating a
proliferative or differentiative disorder in a mammal such as
breast, ovarian or prostate cancer and small cell lung carcinoma
and also major opportunistic infections, immune disorders,
cardiovascular diseases and inflammatory disorders. FBP protein,
analogs, derivatives and their subsequences, anti-FBP antibodies
are also useful in diagnosis of the disorders.

XX
SQ
Sequence 247 AA:

Query Match 58.3%; Score 7; DB 21; Length 247;
Best Local Similarity 100.0%; Pred. No. 2.5; Mismatches 0; Indels 0; Gaps 0;

Qy 6 IIRNGOKL 12

Db 146 IIRNGOKL 152

RESULT 5
AY8304

AY8304 standard; protein; 408 AA.

AC AAB48304;

XX
DT
XX
02-APR-2001 (first entry)

XX
DE
Human ZF23 protein.

XX
S-Phase kinase associated protein; SKP1; SKP2; SKP2-like protein; ZF;
KW CUL-1; cullin; CDC53; p27; cyclin E; Max; Mad; c-MYC; MDM2; p53; Max;
KW Bad; Bcl-2; tumour; cytostatic.

XX
OS
Homo sapiens.

XX
PN WO200075184-A1.

XX
PD
14-DEC-2000.

XX
PR
05-JUN-2000; 2000WO-US15449.

XX
PR
04-JUN-1999; 99US-0137494.

XX
PA
(UWY)
XX
PI Zhang H, Tsvetkov LM, Kondo T;

XX
DR
WPI; 2001-061703/07.

XX
DR
N-PSDB; AAC84616.

XX
PT
Modulating polypeptide levels in a cell, diagnosing and treating tumor,
involves altering levels of proteins such as S-phase kinase associated
proteins 1, 2 and cullin/CDC53 proteins -

XX
PS
Claim 3; Page 139-140; 162pp; English.

XX
 CC The invention relates to methods of altering the polypeptide levels in a
 CC cell, using proteins selected from S-phase kinase associated proteins 1
 CC and 2 (SKP1, SKP2), SKP2-like proteins (2F) and CUL-1 (a member of the
 CC CULIN/CDCC3 family of proteins). The method is useful for altering the
 CC level of p27, cyclin E, Max, Mad, c-MYC, MDM2, p53, Bax, Bad or Bcl-2
 CC polypeptide in a cell, SKP2 and SKP2-like protein levels are useful for
 CC detecting tumours, and in monitoring tumor treatment in a mammal. Agents
 CC that modulate interactions between SKP and target proteins are useful for
 CC treating tumours.

XX
 SQ Sequence 408 AA;

Query	Match	Score	DB	Length
Best	Local	58.38;	22;	408;
Matches	7;	Similarity	100.0%;	Pred. No. 4.1;
Qy	7;	Conservative	0;	Mismatches 0;
Db	141	IHNQKL	12	Indels 0;
				Gaps 0;

RESULT 6
 ID AAY90287
 AC AAY90287;
 DT 24-OCT-2000 (first entry)
 XX DE Human peptidase, HPEP-4 protein sequence.
 XX KW Human; peptidase; cell proliferative disorder; arteriosclerosis;
 KW psoriasis; myelofibrosis; cancer; autoimmune disorder; Crohn's disease;
 KW inflammatory disorder; AIDS; anaemia; allergy; asthma; atherosclerosis;
 KW Gräve's disease; multiple sclerosis; scleroderma; infection; diabetes;
 KW metabolic disorder; Addison's disease; cystic fibrosis; diagnosis;
 KW glycogen storage disease; obesity; therapy; HPEP-4.
 XX OS Homo sapiens.
 PN WO20042201-A2.
 PD 20-JUL-2000.
 XX PF 11-JAN-2000; 2000WO-US00641.
 PR 11-JAN-1999; 99US-0172247.
 PR 03-MAY-1999; 99US-0132253.
 PR 27-MAY-1999; 99US-0136653.
 PA (INCY-) INCYTE PHARM INC.
 PI Bandman O, Hillman JL, Tang YT, Azimzai Y, Baughn MR, Ihal P;
 PI Yue H, Lu DAM;
 DR WPI; 2000-492832/42.
 DR N-PSDB; AAJ37660.

XX An isolated polypeptide for diagnosis, prevention and treatment of
 PT cell proliferative, autoimmune/ inflammatory and metabolic disorders
 PT comprises a sequence encoding a human peptidase -
 PS Claim 2; Page 95-96; 131pp; English.

XX This sequence represents a human peptidase, designated HPEP-4. The
 CC invention relates to 18 human peptidases designated HPEP-1 to HPEP-18,
 CC respectively. The peptidases can be used for treating a disease or
 CC condition associated with decreased expression or over expression of
 CC functional human peptidases. The cell diseases that can be diagnosed,
 CC prevented and treated include cell proliferative disorders (such as
 CC arteriosclerosis, psoriasis, myelofibrosis, and cancers),
 CC autoimmune/inflammatory disorders (such as AIDS, anaemia, allergies,

CC Crohn's disease, asthma, atherosclerosis, Grave's disease, multiple
 CC sclerosis, and scleroderma), infections, and metabolic disorders (such as
 CC Addison's disease, diabetes, cystic fibrosis, glycogen storage diseases
 CC and obesity).

XX SQ Sequence 471 AA;

Query	Match	Score	DB	Length
Best	Local	58.38;	21;	471;
Matches	7;	Similarity	100.0%;	Pred. No. 4.7;
Qy	7;	Conservative	0;	Mismatches 0;
Db	148	IHNQKL	154	Indels 0;
				Gaps 0;

RESULT 7
 ID AAB93481
 AC AAB93481;
 DT 26-JUN-2001 (first entry)
 XX DE Human protein sequence SEQ ID NO:12772.
 XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX OS Homo sapiens.
 PN EP1074617-A2.
 PD 07-FEB-2001.
 XX PF 28-JUL-2000; 2000EP-0116126.
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300233.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-018377.
 PR 09-JUN-2000; 2000JP-0241899.
 XX PA (HELI-) HELIX RES INST.
 XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX DR WPI; 2001-318749/34.
 PR Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX PS Claim 8; SEQ ID 12772; 2537pp + CD ROM; English.
 XX The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5002 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length

CC CDNAS easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13623 to AAH18742 represent human cDNA sequences; AAB2446 to
 CC AAB95833 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 XX of the present invention.

SQ sequence 471 AA;

Query Match 58.3%; Score 7; DB 22; Length 471;
 Best Local Similarity 100.0%; Pred. No. 4.7; Mismatches 0; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 IHNGQKL 12
 XX |||||
 Db 148 IHNGQKL 154

RESULT 8
 AAM47764
 ID AAM47764 standard; Protein: 471 AA.
 XX
 AC AAM47764;
 XX
 XX DT 26-FEB-2002 (first entry)
 XX DE F-cassette structural protein 52.
 XX
 KW F-cassette structural protein 52; tumour; haemopathy; HIV infection;
 KW immunological disease; inflammation; gene therapy; cytostatic;
 KW haemostatic; virucide; immunomodulatory; antiinflammatory.
 XX OS Unidentified.
 XX
 XX CN1305985-A.
 XX
 PD 08-AUG-2001.
 XX
 XX PF 26-JAN-2000; 2000CN-0111548.
 XX
 PR 26-JAN-2000; 2000CN-0111548.
 XX
 PA (BODA-) BODAO GENE TECH CO LTD SHANGHAI.
 XX
 PI Mao Y, Xie Y;
 XX
 DR WRI; 2002-026870/04.
 DR N-PSDB; ABA04771.
 XX
 PT Polypeptide-F-cassette structural protein 52 and polynucleotide for
 PT coding said polypeptide.
 XX
 PS Claim 1; Page 26 (Disclosure); 34pp; Chinese.
 XX
 CC The present sequence is the protein sequence for F-cassette structural
 CC protein 52. The protein and its coding sequence are useful in treating
 CC various diseases, such as malignant tumours, haemopathy, HIV infection,
 CC immunological diseases and inflammations.
 XX
 Sequence 471 AA;

Query Match 58.3%; Score 7; DB 23; Length 471;
 Best Local Similarity 100.0%; Pred. No. 4.7; Mismatches 0; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 IHNGQKL 12
 XX |||||
 Db 148 IHNGQKL 154

RESULT 9
 ABB38678
 ID ABB38678 standard; Protein: 109 AA.
 XX

AC ABP38678;
 XX
 ID AAG37863 standard; Protein: 130 AA.
 XX
 AC AAG37863;
 XX
 AC AAG37863;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 46628.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.

RESULT 10
 AAG37863
 ID AAG37863 standard; Protein: 130 AA.
 XX
 AC AAG37863;
 XX
 AC AAG37863;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 46628.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.

PD	05-SEP-2000.	PR	06-JUL-1999;	9905-0142390.	
XX		PR	08-JUL-1999;	9905-0142803.	
PF	21-FEB-2000;	2000EP-0301439.	PR	09-JUL-1999;	9905-0142920.
XX			PR	12-JUL-1999;	9905-0142977.
PR	25-FEB-1999;	9905-0121825.	PR	13-JUL-1999;	9905-0143542.
PR	03-MAR-1999;	9905-0123180.	PR	14-JUL-1999;	9905-0143624.
PR	03-MAR-1999;	9905-0123548.	PR	15-JUL-1999;	9905-0144005.
PR	23-MAR-1999;	9905-0125788.	PR	16-JUL-1999;	9905-0144085.
PR	25-MAR-1999;	9905-0126264.	PR	16-JUL-1999;	9905-0144335.
PR	29-MAR-1999;	9905-0126785.	PR	19-JUL-1999;	9905-0144325.
PR	01-APR-1999;	9905-0127462.	PR	20-JUL-1999;	9905-0144632.
PR	01-APR-1999;	9905-0128234.	PR	19-JUL-1999;	9905-0144332.
PR	05-APR-1999;	9905-0128714.	PR	19-JUL-1999;	9905-014433.
PR	16-APR-1999;	9905-0129845.	PR	19-JUL-1999;	9905-0144334.
PR	19-APR-1999;	9905-0130077.	PR	19-JUL-1999;	9905-0144335.
PR	21-APR-1999;	9905-0130449.	PR	20-JUL-1999;	9905-0144505.
PR	21-APR-1999;	9905-0130510.	PR	20-JUL-1999;	9905-01445087.
PR	21-APR-1999;	9905-0130891.	PR	20-JUL-1999;	9905-0144884.
PR	06-MAY-1999;	9905-0132486.	PR	21-JUL-1999;	9905-0144814.
PR	07-MAY-1999;	9905-0132863.	PR	21-JUL-1999;	9905-0145086.
PR	11-MAY-1999;	9905-0134256.	PR	23-JUL-1999;	9905-0145088.
PR	04-MAY-1999;	9905-0132484.	PR	22-JUL-1999;	9905-0145085.
PR	05-MAY-1999;	9905-0132485.	PR	22-JUL-1999;	9905-0145087.
PR	14-MAY-1999;	9905-0132486.	PR	22-JUL-1999;	9905-0145089.
PR	14-MAY-1999;	9905-0132863.	PR	22-JUL-1999;	9905-0145192.
PR	18-MAY-1999;	9905-0134768.	PR	23-JUL-1999;	9905-0145145.
PR	19-MAY-1999;	9905-0134941.	PR	23-JUL-1999;	9905-0145218.
PR	14-MAY-1999;	9905-0134218.	PR	23-JUL-1999;	9905-0145224.
PR	14-MAY-1999;	9905-0134219.	PR	26-JUL-1999;	9905-0145276.
PR	24-MAY-1999;	9905-0134221.	PR	27-JUL-1999;	9905-0145913.
PR	14-MAY-1999;	9905-0134370.	PR	27-JUL-1999;	9905-0145918.
PR	06-MAY-1999;	9905-0134370.	PR	27-JUL-1999;	9905-0145919.
PR	20-MAY-1999;	9905-0134370.	PR	28-JUL-1999;	9905-0145951.
PR	20-MAY-1999;	9905-0135124.	PR	23-JUL-1999;	9905-0145386.
PR	21-MAY-1999;	9905-0135353.	PR	02-AUG-1999;	9905-0146388.
PR	24-MAY-1999;	9905-0135629.	PR	02-AUG-1999;	9905-0146389.
PR	25-MAY-1999;	9905-0136021.	PR	03-AUG-1999;	9905-0147038.
PR	27-MAY-1999;	9905-0136392.	PR	04-AUG-1999;	9905-0147204.
PR	28-MAY-1999;	9905-0136782.	PR	04-AUG-1999;	9905-0147302.
PR	01-JUN-1999;	9905-0137222.	PR	05-AUG-1999;	9905-0147192.
PR	03-JUN-1999;	9905-0137528.	PR	05-AUG-1999;	9905-0147260.
PR	04-JUN-1999;	9905-0137502.	PR	06-AUG-1999;	9905-0147303.
PR	07-JUN-1999;	9905-0137724.	PR	06-AUG-1999;	9905-0147416.
PR	08-JUN-1999;	9905-0138094.	PR	09-AUG-1999;	9905-0147493.
PR	10-JUN-1999;	9905-0138540.	PR	09-AUG-1999;	9905-0147735.
PR	10-JUN-1999;	9905-0138847.	PR	10-AUG-1999;	9905-0148171.
PR	14-JUN-1999;	9905-0139119.	PR	11-AUG-1999;	9905-0148319.
PR	16-JUN-1999;	9905-0139452.	PR	12-AUG-1999;	9905-0148341.
PR	16-JUN-1999;	9905-0139492.	PR	13-AUG-1999;	9905-0148655.
PR	18-JUN-1999;	9905-0139454.	PR	16-AUG-1999;	9905-0148684.
PR	18-JUN-1999;	9905-0139455.	PR	17-AUG-1999;	9905-0149368.
PR	18-JUN-1999;	9905-0139456.	PR	18-AUG-1999;	9905-0149426.
PR	18-JUN-1999;	9905-0139457.	PR	20-AUG-1999;	9905-0149722.
PR	18-JUN-1999;	9905-0139458.	PR	20-AUG-1999;	9905-0149723.
PR	18-JUN-1999;	9905-0139459.	PR	23-AUG-1999;	9905-0149929.
PR	18-JUN-1999;	9905-0139460.	PR	23-AUG-1999;	9905-0149930.
PR	18-JUN-1999;	9905-0139461.	PR	25-AUG-1999;	9905-0150566.
PR	18-JUN-1999;	9905-0139462.	PR	26-AUG-1999;	9905-0150884.
PR	18-JUN-1999;	9905-0139463.	PR	27-AUG-1999;	9905-0151065.
PR	18-JUN-1999;	9905-0139750.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139763.	PR	27-AUG-1999;	9905-0151080.
PR	21-JUN-1999;	9905-0139817.	PR	30-AUG-1999;	9905-0151303.
PR	22-JUN-1999;	9905-0139889.	PR	31-AUG-1999;	9905-0151303.
PR	23-JUN-1999;	9905-0140353.	PR	01-SEP-1999;	9905-0151930.
PR	23-JUN-1999;	9905-0140354.	PR	07-SEP-1999;	9905-0152363.
PR	24-JUN-1999;	9905-0140695.	PR	10-SEP-1999;	9905-0153070.
PR	28-JUN-1999;	9905-0140823.	PR	13-SEP-1999;	9905-0153758.
PR	30-JUN-1999;	9905-0140991.	PR	15-SEP-1999;	9905-0154018.
PR	01-JUL-1999;	9905-0141287.	PR	16-SEP-1999;	9905-0154039.
PR	02-JUL-1999;	9905-0141842.	PR	20-SEP-1999;	9905-0154779.
PR		9905-0142154.	PR	22-SEP-1999;	9905-0155139.

PR	24-SEP-1999;	990S-0155658.
PR	28-SEP-1999;	990S-0155658.
PR	29-SEP-1999;	990S-0156595.
PR	04-OCT-1999;	990S-0157117.
PR	05-OCT-1999;	990S-0157753.
PR	06-OCT-1999;	990S-0157855.
PR	07-OCT-1999;	990S-0158029.
PR	08-OCT-1999;	990S-0158232.
PR	12-OCT-1999;	990S-0153659.
PR	13-OCT-1999;	990S-0159293.
PR	13-OCT-1999;	990S-0159294.
PS	13-OCT-1999;	990S-0159295.
XX	14-OCT-1999;	990S-0159329.
CC	14-OCT-1999;	990S-0159330.
CC	14-OCT-1999;	990S-015331.
PR	14-OCT-1999;	990S-0159637.
PR	14-OCT-1999;	990S-0159638.
PR	18-OCT-1999;	990S-015584.
PR	21-OCT-1999;	990S-0160741.
PR	21-OCT-1999;	990S-0160767.
PR	21-OCT-1999;	990S-0160768.
PR	21-OCT-1999;	990S-0160770.
PR	21-OCT-1999;	990S-0160814.
PR	21-OCT-1999;	990S-0160815.
PR	22-OCT-1999;	990S-0160980.
PR	22-OCT-1999;	990S-0160981.
PR	22-OCT-1999;	990S-0160989.
PR	25-OCT-1999;	990S-0161404.
PR	25-OCT-1999;	990S-0161405.
PR	25-OCT-1999;	990S-0161406.
PR	26-OCT-1999;	990S-0161359.
PR	26-OCT-1999;	990S-0161360.
PR	26-OCT-1999;	990S-0161361.
PR	28-OCT-1999;	990S-0161920.
PR	28-OCT-1999;	990S-0161992.
PR	28-OCT-1999;	990S-0161993.
PR	29-OCT-1999;	990S-0162142.
Query Match	50.0%; Score 6; DB 21; Length 130;	
Best Local Similarity	100.0%; Pred. No. 18;	
Matches	6; Conservative 0; Mismatches 0;	
Qy	6 IHNGQK 11	
Db	14 IHNGQK 19	
RESULT 11		
ABG24051		
ID	ABG24051 standard; Protein: 162 AA.	
XX		
AC	ABG24051;	
XX		
18-FEB-2002	(first entry)	
DE	Novel human diagnostic protein #24042.	
XX		
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.	
XX		
OS	Homo sapiens.	
XX		
PN	W0200175067-A2.	
XX		
PD	11-OCT-2001.	
XX		
PF	30-MAR-2001; 2001WO-US08631.	
XX		
PR	31-MAR-2000; 2000US-0540217.	
PR	23-AUG-2000; 2000US-0549167.	
XX		
XX		
(HYSEQ -) HYSEQ INC.		
PI	Doucette-Stamm LA, Bush D;	
PR	23-SEP-1999;	990S-0155658.
PR	28-SEP-1999;	990S-0155658.
PR	29-SEP-1999;	990S-0156595.
PR	04-OCT-1999;	990S-0157117.
PR	05-OCT-1999;	990S-0157753.
PR	06-OCT-1999;	990S-0157855.
PR	07-OCT-1999;	990S-0158029.
PR	08-OCT-1999;	990S-0158232.
PR	12-OCT-1999;	990S-0153659.
PR	13-OCT-1999;	990S-0159293.
PR	13-OCT-1999;	990S-0159294.
PS	13-OCT-1999;	990S-0159295.
XX	14-OCT-1999;	990S-0159329.
CC	14-OCT-1999;	990S-0159330.
CC	14-OCT-1999;	990S-015331.
PR	14-OCT-1999;	990S-0159637.
PR	14-OCT-1999;	990S-0159638.
PR	18-OCT-1999;	990S-015584.
PR	21-OCT-1999;	990S-0160741.
PR	21-OCT-1999;	990S-0160767.
PR	21-OCT-1999;	990S-0160768.
PR	21-OCT-1999;	990S-0160770.
PR	21-OCT-1999;	990S-0160814.
PR	21-OCT-1999;	990S-0160815.
PR	22-OCT-1999;	990S-0160980.
PR	22-OCT-1999;	990S-0160981.
PR	22-OCT-1999;	990S-0160989.
PR	25-OCT-1999;	990S-0161404.
PR	25-OCT-1999;	990S-0161405.
PR	25-OCT-1999;	990S-0161406.
PR	26-OCT-1999;	990S-0161359.
PR	26-OCT-1999;	990S-0161360.
PR	26-OCT-1999;	990S-0161361.
PR	28-OCT-1999;	990S-0161920.
PR	28-OCT-1999;	990S-0161992.
PR	28-OCT-1999;	990S-0161993.
PR	29-OCT-1999;	990S-0162142.
Query Match	50.0%; Score 6; DB 21; Length 130;	
Best Local Similarity	100.0%; Pred. No. 18;	
Matches	6; Conservative 0; Mismatches 0;	
Qy	1 TSPLNI 6	
Db	52 TSPLNI 57	
RESULT 12		
ABP38390		
ID	ABP38390 standard; Protein: 467 AA.	
XX		
AC	ABP38390;	
XX		
DT	24-JUL-2002 (first entry)	
XX		
DE	Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3235.	
XX		
KW	Staphylococcus epidermidis; open reading frame; ORF; bacterial infection; antibacterial; gene therapy.	
XX		
OS	Staphylococcus epidermidis.	
XX		
PN	US380370-B1.	
XX		
PD	30-APR-2002.	
XX		
PF	13-AUG-1998; 980S-0134001.	
XX		
PR	14-AUG-1997; 970S-055779P.	
PR	08-NOV-1997; 970S-064964P.	
XX		
PA	(GENO-) GENOME THERAPEUTICS CORP.	
XX		

XX WPI; 2002-381255/41.
 DR DR
 XX N-PSDB; AHN0935.
 PT Novel isolated nucleic acid encoding a *Staphylococcus epidermidis* -
 PT polypeptide, useful for diagnosing and treating bacterial infections -
 XX disclosure; SEQ ID 3325; 267pp; English.
 XX ABN90338 to ABN93374 represent *Staphylococcus epidermidis* open reading
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
 CC given in ABP3124 to ABP31960. The *S. epidermidis* sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences
 CC can also be used in the diagnosis and treatment of bacterial infections,
 CC particularly *S. epidermidis* infections. The sequences can be used to
 CC screen for compounds able to interfere with the *S. epidermidis* life
 CC cycle or inhibit *S. epidermidis* infection.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site.
 XX SQ Sequence 467 AA;

Query Match 50.0%; Score 6; DB 23; Length 467;
 Best local similarity 100.0%; Pred. No. 63;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 NIHNGQ 10
 Db 456 NIHNGQ 461

RESULT 13
 AAG82404
 ID AAG82404 standard; Protein: 472 AA.

XX AC AAG82404;
 DT 03-SEP-2001 (first entry)

DE S. epidermidis open reading frame protein sequence SEQ ID NO:1902.
 XX KW *Staphylococcus epidermidis* SRL strain; infection; diagnosis;
 KW vaccination; endocarditis.
 XX OS *Staphylococcus epidermidis*.
 XX PN WO200134809-A2.

XX PD 17-MAY-2001.
 XX PF 09-NOV-2000; 2000WO-US30782.

XX PR 09-NOV-1999; 99US-0164258.

XX PA (GLAXO) GLAXO GROUP LTD.
 PT Kimmerly WJ;
 PT XX

XX DR WPI; 2001-316495/33.
 XX DR N-PSDB; AHN53254.

PT Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,
 PT useful for vaccinating against infections, e.g. endocarditis -
 XX Claim 18; Page 519-520; 218pp; English.
 XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC *S. epidermidis* polypeptides (II) via the production of vectors
 CC containing them which are used to produce host cells which express the

CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to
 CC AAH55030 represent specifically claims 5. *S. epidermidis* genomic DNA
 CC polynucleotide sequences from the present invention. AAH55091 to
 CC AAH55098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.
 XX SQ Sequence 472 AA;

Query Match 50.0%; Score 6; DB 22; Length 472;
 Best local similarity 100.0%; Pred. No. 64;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 NIHNGQ 10
 Db 461 NIHNGQ 466

RESULT 14
 AAW13187
 ID AAW13187 standard; Protein: 20 AA.

XX AC AAW13187;

XX DT 19-MAY-1997 (first entry)

DE Fragment of p53 binding protein WBPI.

XX KW p53; binding protein; WBPI; antibody; diagnosis; treatment;
 KW neoplastic; preneoplastic; disease; agonist; antagonist;
 KW augmentation; inhibition; complex formation; neoplasia; apoptosis;
 KW reperfusion; injury; myocardial infarction; stroke; AIDS;
 KW traumatic brain; neurodegenerative; aging; ischaemia; toxæmia;
 KW infection; hepatitis; probe; genetic; forensic identification;
 KW fragment.
 XX OS Homo sapiens.
 XX PN W09514777-A1.
 XX PD 01-JUN-1995.
 XX PF 21-NOV-1994; 94WO-US13499.
 XX PR 22-NOV-1993; 93US-0156571.

XX PA (ONYX-) ONYX PHARM.

XX PI Bischoff JR, Wu L;

XX DR WPI; 1995-206934/27.

XX PT New p53-binding polypeptide(s) WBPI and p53uIC - used to develop
 PT prods. for screening assays and for use in diagnosis and therapy of
 PR diseases, esp. neoplasia
 XX Claim 8; Page 38; 90pp; English.

XX The present sequence is a fragment of the p53 binding protein WBPI,
 CC which can be used to generate antibodies for the diagnosis of
 CC (pre)neoplastic diseases, and WBPI (anti)agonists. The (anti)agonists
 CC inhibit neoplasia or apoptosis, useful in the treatment of, e.g.
 CC reperfusion injury, myocardial infarction, stroke, traumatic brain
 CC injury, neurodegenerative diseases, aging, ischaemia, toxæmia,

THIS PAGE BLANK (USPTO)

CC infection, AIDS and hepatitis. The WBPL encoding cDNA, which was isolated from a HeLa cell derived cDNA library using the yeast two-hybrid system, can be used to generate probes for the diagnosis of (pre)neoplastic pathological conditions and genetic diseases, and the forensic identification of human individuals.

SQ Sequence 20 AA;

Query Match 41.7%; Score 5; DB 16; Length 20;

Best Local Similarity 100.0%; Prd. No. 41; Mismatches 5; Matches 5; Conservative 0; Indels 0; Gaps 0;

QY 1 TSPN 5
15 TSPLN 19

RESULT 15
AAW57358
ID AAW57358
XX standard; peptide; 20 AA.

AC AAW57358;
XX
DT 11-AUG-1998 (first entry)

DE Human WBPL immunogenic peptide SEQ ID NO:90.
XX

KW Human; WBPL; p53; cancer; interacting protein; screen; diagnosis; genetic disease; forensic identification; nuclear phosphoprotein; cellular proliferation; neoplastic transformation; p53UBC; immunogen.

KW XX
OS Homo sapiens.
XX
PN US756669-A.
XX
PD 26-MAY-1998.
XX
PF 01-MAR-1995; 95US-0399696.
XX
PR 02-MAR-1995; 95US-0399696.
PR 21-NOV-1993; 93US-0156571.
PR 21-NOV-1994; 94WO-US13499.
XX
PA (ONYX-) CONYX PHARM INC.
XX
PT Bischoff, JR, Wu L;
XX
DR WPI; 1998-321626/28.

XX
PT WBPL, p53-interacting protein - useful screening agents for treatments of p53 related cancers
XX
PS Disclosure; Column 24; 68pp; English.

XX
CC The present sequence represents an immunogenic peptide from WBPL (cellular proliferation and neoplastic transformation) polypeptide, which can be used to screen bacteriophage antibody display libraries or to immunize a rabbit. WBPL is a p53-interacting protein that can be used to screen for agents for diagnosis or treatment of cancer and genetic diseases and for forensic identification of human individuals. The p53 protein is a nuclear phosphoprotein involved in control of cellular proliferation, and mutations in it are associated with human cancers.
XX
SQ Sequence 20 AA;
Query Match 41.7%; Score 5; DB 19; Length 20;
Best Local Similarity 100.0%; Prd. No. 41; Mismatches 5; Matches 5; Conservative 0; Indels 0; Gaps 0;

Search completed: January 3, 2003, 14:27:45
Job time : 37 secs

QY 1 TSPN 5
15 TSPLN 19